

THIS WEEK

EDITORIALS

HERITAGE The true value of cultural sites to science and beyond **p.302**

WORLD VIEW The climate needs protest and civil disobedience **p.303**



OPEN WIDE Oral bumps help alligators and crocodiles find food **p.304**

America's carbon compromise

As looming tax increases and budget cuts threaten to plunge the US economy back into recession, Congress should take a hard look at introducing a carbon tax as an important part of the solution.

This week, a reinvigorated Barack Obama returned to the White House knowing that he was poised on the edge of a fiscal cliff. Rather than relishing his victory last week, Obama must immediately set about crafting a compromise on deficit reduction with congressional leaders. The stakes could hardly be higher — for science, for US citizens and, indeed, for the world. In the event of failure, a budgetary time-bomb of tax increases and sweeping budget cuts will detonate on 2 January. As well as resulting in indiscriminate cuts to funds for scientific research and many other areas, it could knock the United States back into recession and deliver yet another blow to an already fragile global economy.

Faced with such dire consequences, one might expect that all the financial options would be on the table, especially the good ones. Unfortunately, this is not the case, at least not yet.

So far, lawmakers have rehashed long-standing disputes about the size of government and the social safety net, but have ignored ideas that could transform the fiscal challenge into an opportunity. One such proposal is the carbon tax, which could bring financial and political benefits for all and chart a new course forward for energy independence and global warming (see page 309). It is a solution that is every bit as improbable as it is logical, but one should remember Winston Churchill's assessment of the United States' tendency to do the right thing — once all the alternatives have been exhausted.

Just consider the possibilities. To put a levy on carbon would raise revenues that could be used to offset lower tax rates for individuals and businesses. This is what conservatives say they want to do. It would put more income — and thus choice — in the hands of consumers. Economists like the idea for more fundamental reasons. Generally, it is best to tax things that one wishes to discourage (such as smoking) rather than those that should be encouraged (such as working). Environmentalists like the idea of a carbon tax because it could generate some much-needed revenue for clean-energy research and development while reducing carbon emissions.

The numbers are not negligible. An analysis conducted in August by economists at the Massachusetts Institute of Technology (MIT) in Cambridge showed that a carbon tax of US\$20 per tonne of carbon dioxide from fossil fuels, if instituted in 2013 and increased by 4% per year, would raise \$1.5 trillion over the course of a decade. Averaged out, this amounts to \$150 billion annually — a sizeable chunk of the trillion-dollar deficits that the US government has been running in recent years. Scholars at the Brookings Institution, a centrist think tank in Washington DC, advocate ramping federal investments in energy research up from \$3.8 billion now to \$30 billion annually, to drive down the cost of low-carbon energy (including cleaner-burning coal). It is an ambitious proposal, and would leave a pile of cash that could be redistributed elsewhere for beneficial use.

Conservatives loathe taxes, and US politicians obsess over energy prices, but a revenue-neutral carbon tax would get around these

problems. The MIT analysis found that the economy benefited regardless of whether the money was reinvested in social programmes or redistributed in the form of lower taxes and cash payments to offset higher energy costs for the poor. For environmentalists, the problem with a carbon tax is that it does not technically limit emissions, but the MIT model suggests that it would perform quite well: carbon emissions fall to 14% below 2006 levels by 2020 as consumers and businesses find ways to reduce their energy use in response to higher prices.

"A carbon tax would depend on political courage and a break with party orthodoxy."

Opposition to the idea may not be what it was. For example, on 13 November, the American Enterprise Institute hosted a conference in Washington DC on the economics of a carbon tax. The institute is a conservative think tank, and its officials have previously raised doubts about climate science. The idea has also been bubbling up in other right-leaning think tanks as a conservative solution to reduce greenhouse gases.

The problem is that to enact a carbon tax would depend on political courage and a willingness to break with party orthodoxy, rare traits in Washington today. President Obama has made energy and climate part of his agenda for the second term, but his first — and perhaps biggest — opportunity to make good on that promise will come in the next few weeks. As US politicians contemplate diving into the fiscal abyss, they would be wise to consider a painless policy that benefits everyone. ■

A shaky restart

Japan still has lessons to learn from Fukushima if it is to convince the public about nuclear energy.

The nuclear disaster that followed the March 2011 tsunami in Japan uncovered serious flaws in the country's nuclear-safety regulations. Japan learned its lesson: it started putting a premium on safety, and is doing everything it can to assure a wary public that similar mistakes will not be made again. Well, that was the hope. Two recent revelations show that it could still do much more.

The country's Nuclear Regulation Authority (NRA) was set up to right the wrongs of the previous regulatory infrastructure. One of its first tasks was to draw up new safety standards for reactor operations. The NRA formed an investigation team of six experts, which held its first meeting on 25 October. The team is expected to submit its report in time for the NRA to put the standards up for public comment in the spring and to make them law in July 2013.

Last week, Japanese media reported that four of those experts have

received regular stipends or one-time grants from the nuclear industry. Nuclear engineer Akio Yamamoto of Nagoya University, for example, has received at least ¥50,000 (US\$630) over the past three years from each of three companies related to nuclear energy, including Nuclear Engineering, a firm in Osaka that is affiliated with Kansai Electric Power. Although there is no suggestion that Yamamoto has done anything wrong, he also received some ¥27 million in grants from eight nuclear-energy companies during that period, as well as an undisclosed amount from Mitsubishi Heavy Industries, which builds reactors.

An NRA spokesman defended the team's composition, arguing that the report will be used only as a reference for the five NRA commissioners who will ultimately decide on the policy. (Those commissioners do not have similar ties with industry.) If the NRA had tried to rule out everyone with any connection to the industry when choosing the experts, the spokesman said, it would have run out of people.

These are fair points, and the fact that the team members had to disclose their contributions at all was a laudable nod to transparency. But playing down the importance of the report by saying it will just be used as a reference is unconvincing.

Much of the uproar over the handling of the Fukushima disaster came from the public perception that conflicts of interests led regulators, who were too tightly tied to the nuclear industry, to favour cost-savings over safety. The NRA, created in large part as an answer to that criticism, has itself been lambasted for moving many staff from the old regulatory structure to the new organization, including its head, Shunichi Tanaka. It seems that Japanese policy-makers, despite the many public demonstrations, still haven't got to grips with the tendency for conflicts of

interest to lead to bad decisions and, even if they don't, to breed mistrust.

Similarly troubling is the rush with which the government reopened two of the country's shuttered nuclear reactors in July without fully evaluating the seismology of the ground beneath. Last week, at its second meeting, a subcommittee of the NRA could not confirm whether

"Japan was supposed to emerge with a new respect for reactor safety."

a fault line running under a seawater-intake channel — used to cool the reactors in an emergency — is active.

At stake is whether the fault is a landslide fault or a more dangerous, deeper tectonic one. The NRA has ordered Kansai Electric, the plant's operator, to dig trenches to investigate the geology more thoroughly. That should take less than two months, but existing facilities at the plant are in the way, making it much more complicated — and expensive.

Even if the risk from that fault is trivial, as many think, critics point out that the threat of shaking from nearby faults, the potential size of a tsunami and the possibility of structural defects like those found at Fukushima have not been adequately characterized.

Large sectors of the public opposed the reactor restarts with demonstrations of a fervour not seen in Japan in decades. The country had already proved that it could get by, at least in the short term, with no nuclear power. Some scientists had pointed out the uncertainty over the seismic fault, and suggested how to deal with it, before the reactors were restarted. Japan was supposed to emerge from the Fukushima crisis with a new respect for reactor safety and better awareness of the need to convince people of that safety. It hasn't made a very good start. ■

Save scientific sites

The push to conserve cultural-heritage sites must not leave out areas of interest to science.

It is possible for an outsider to visit and enjoy the ancient city of Samarkand in Uzbekistan, home to the fifteenth-century observatory of the astronomer Ulugh Beg. That is in part thanks to the United Nations Educational, Scientific and Cultural Organization (UNESCO) World Heritage Convention, which this week celebrates its 40th anniversary (see page 328). Samarkand was put on the World Heritage List in 2001; the listing gives it important protection from the ongoing political chaos that has followed the collapse of the Soviet Union.

The observatory must have been lovely during the two decades that it was active. Contemporary reports describe splendid architecture and exquisite tiling and mosaics. Frescos illustrating the orbits of the planets and the exact positions of stars adorned the observatory's inner walls. It was largely destroyed by God-fearing hordes in 1449, but the innovative work of its scientists survived to influence Western astronomy and algebra. Using a sextant 40 metres in radius, astronomers at the observatory recalculated the positions of nearly 1,000 stars and compiled their results in the widely translated 1437 star catalogue *Zij-i Sultani*. They stabilized the sextant by anchoring it in a 2-metre-wide trench dug into a hill in the plane of the meridian. The measurements were of unprecedented accuracy: the astronomers used them to recalculate trigonometric tables and to calculate the length of the sidereal year (the time taken for Earth to orbit the Sun once in relation to the fixed stars) to within one minute of the measurement now accepted. Archaeologists discovered the remains of the observatory in 1908.

A place on the World Heritage List means that a site must be maintained with international-standard conservation methods, and not spoiled with inappropriate building development. International inspection teams visit to ensure compliance. Inspections of Samarkand revealed in the mid-2000s that conservation was under par, and

that city planners, supported by local politicians, alarmingly failed to respect the integrity of the site. UNESCO called for increased monitoring of the site and threatened to place it on the List of World Heritage in Danger — although it has so far avoided this designation.

In the past few decades, scientists from many disciplines have developed techniques — ranging from lasers to nuclear technologies and microbiology — for conserving and restoring artworks and monuments. Scientists at the Foundation for Research and Technology — Hellas in Heraklion, Greece, for example, invented a laser with one beam in the ultraviolet range and another in the infrared to clean a frieze on the Parthenon, part of a World Heritage Site, without damaging its surface. The widely publicized work on the frieze was completed in 2005.

Understandably, scientists would like more funding to allow them to fine-tune such techniques. That is hard to justify generically — each archaeological site or monument has its own problems, with technical solutions that must be worked out on an individual basis. The story of Samarkand shows that politics — and thus the World Heritage List — is at least as important as science to the conservation of important monuments around the world. Funds for cultural-heritage technologies must be maintained as part of a broad approach to consider cultural heritage more widely. City and regional plans to cope with climate change, for example, should be required to consider the impact on cultural heritage. The European Commission is quietly voicing support for such a push, and it should be encouraged to speak louder.

Very few of the 962 entries on the World Heritage List involve scientific sites, perhaps because science is not automatically thought of as a part of culture. However, astronomers have begun to do something about this. In the 2009 International Year of Astronomy, astronomers worked with a UNESCO advisory group, the International Council on Monuments and Sites, to produce a list of astronomical sites that they think are, like the Samarkand Observatory, worth saving. These include the nineteenth-century Royal Observatory in Cape Town,

South Africa. If this site is designated as a World Heritage Site, interest surrounding it might encourage the much-desired development of science in the country. Other scientists should follow the astronomers' example. ■

➔ **NATURE.COM**
To comment online,
click on Editorials at:
go.nature.com/xhbnq



Be persuasive. Be brave. Be arrested (if necessary)

A resource crisis exacerbated by global warming is looming, argues financier Jeremy Grantham. More scientists must speak out.

I have yet to meet a climate scientist who does not believe that global warming is a worse problem than they thought a few years ago. The seriousness of this change is not appreciated by politicians and the public. The scientific world carefully measures the speed with which we approach the cliff and will, no doubt, carefully measure our rate of fall. But it is not doing enough to stop it. I am a specialist in investment bubbles, not climate science. But the effects of climate change can only exacerbate the ecological trouble I see reflected in the financial markets — soaring commodity prices and impending shortages.

My firm warned of vastly inflated Japanese equities in 1989 — the grandmother of all bubbles — US growth stocks in 2000 and everything risky in late 2007. The usual mix of investor wishful thinking and dangerous and cynical encouragement from industrial vested interests made these bubbles possible. Prices of global raw materials are now rising fast. This does not constitute a bubble, however, but is a genuine paradigm shift, perhaps the most important economic change since the Industrial Revolution. Simply, we are running out.

The price index of 33 important commodities declined by 70% over the 100 years up to 2002 — an enormous help to industrialized countries in getting rich. Only one commodity, oil, had been flat until 1972 and then, with the advent of the Organization of the Petroleum Exporting Countries, it began to rise. But since 2002, prices of almost all the other commodities, plus oil, tripled in six years; all without a world war and without much comment. Even if prices fell tomorrow by 20% they would still on average have doubled in 10 years, the equivalent of a 7% annual rise.

This price surge is a response to global population growth and the explosion of capital spending in China. Especially dangerous to social stability and human well-being are food prices and food costs. Growth in the productivity of grains has fallen to 1.2% a year, which is exactly equal to the global population growth rate. There is now no safety margin.

Then there is the impending shortage of two fertilizers: phosphorus (phosphate) and potassium (potash). These two elements cannot be made, cannot be substituted, are necessary to grow all life forms, and are mined and depleted. It's a scary set of statements. Former Soviet states and Canada have more than 70% of the potash. Morocco has 85% of all high-grade phosphates. It is the most important quasi-monopoly in economic history.

What happens when these fertilizers run out is a question I can't get satisfactorily answered and, believe me, I have tried. There seems to be only one conclusion: their use must be drastically reduced in the next 20–40 years or we will begin to starve.

The world's blind spot when it comes to the

fertilizer problem is seen also in the shocking lack of awareness on the part of governments and the public of the increasing damage to agriculture by climate change; for example, runs of extreme weather that have slashed grain harvests in the past few years. Recognition of the facts is delayed by the frankly brilliant propaganda and obfuscation delivered by energy interests that virtually own the US Congress. (It is not unlike the part played by the financial industry when investment bubbles start to form ... but that, at least, is only money.) We need oil producers to leave 80% of proven reserves untapped to achieve a stable climate. As a former oil analyst, I can easily calculate oil companies' enthusiasm to leave 80% of their value in the ground — absolutely nil.

The damaging effects of climate change are accelerating. James

Hansen of NASA has screamed warnings for 30 years. Although at first he was dismissed as a madman, almost all his early predictions, disturbingly, have proved conservative in relation to what has actually happened. In 2011, Hansen was arrested in Washington DC, alongside Gus Speth, the retired dean of Yale University's environmental school; Bill McKibben, one of the earliest and most passionate environmentalists to warn about global warming; and my daughter-in-law, all for protesting over a pipeline planned to carry Canadian bitumen to refineries in the United States, bitumen so thick it needs masses of water even to move it. From his seat in jail, Speth said that he had held some important positions in Washington, but none more important than this one.

President Barack Obama missed the chance of a lifetime to get a climate bill passed, and his great environmental and energy scientists John

Holdren and Steven Chu went missing in action. Scientists are understandably protective of the dignity of science and are horrified by publicity and overstatement. These fears, unfortunately, are not shared by their opponents, which makes for a rather painful one-sided battle. Overstatement may generally be dangerous in science (it certainly is for careers) but for climate change, uniquely, understatement is even riskier and therefore, arguably, unethical.

It is crucial that scientists take more career risks and sound a more realistic, more desperate, note on the global-warming problem. Younger scientists are obsessed by thoughts of tenure, so it is probably up to older, senior and retired scientists to do the heavy lifting. Be arrested if necessary. This is not only the crisis of your lives — it is also the crisis of our species' existence. I implore you to be brave. ■

Jeremy Grantham is co-founder and chief investment strategist at GMO, and co-chair of the Grantham Foundation for the Protection of the Environment, in Boston, Massachusetts.
e-mail: jeremy@granthamfoundation.org

IT IS CRUCIAL THAT
SCIENTISTS
SOUND A MORE
REALISTIC,
MORE
DESPERATE,
NOTE ON GLOBAL
WARMING.

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/k8mrbe

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

PHYSIOLOGY

How cold triggers fat formation

In frigid temperatures, mice ramp up the production of heat-generating brown fat by decreasing the levels of a small RNA molecule.

Because brown fat burns energy — unlike its unpopular cousin, white fat — its production is an attractive target for obesity and diabetes therapies. Markus Stoffel at the Swiss Federal Institute of Technology in Zurich and his colleagues found that exposure to cold reduced the expression of microRNA-133 in brown and subcutaneous white fat. Inhibiting miRNA-133 promoted brown-fat formation, whereas forcing miRNA-133 expression switched off brown-fat production.

The small RNA acted by directly inhibiting PRDM16, a protein that is central to the production of brown fat from white-fat-cell precursors. *Nature Cell Biol.* <http://dx.doi.org/10.1038/ncb2612> (2012)

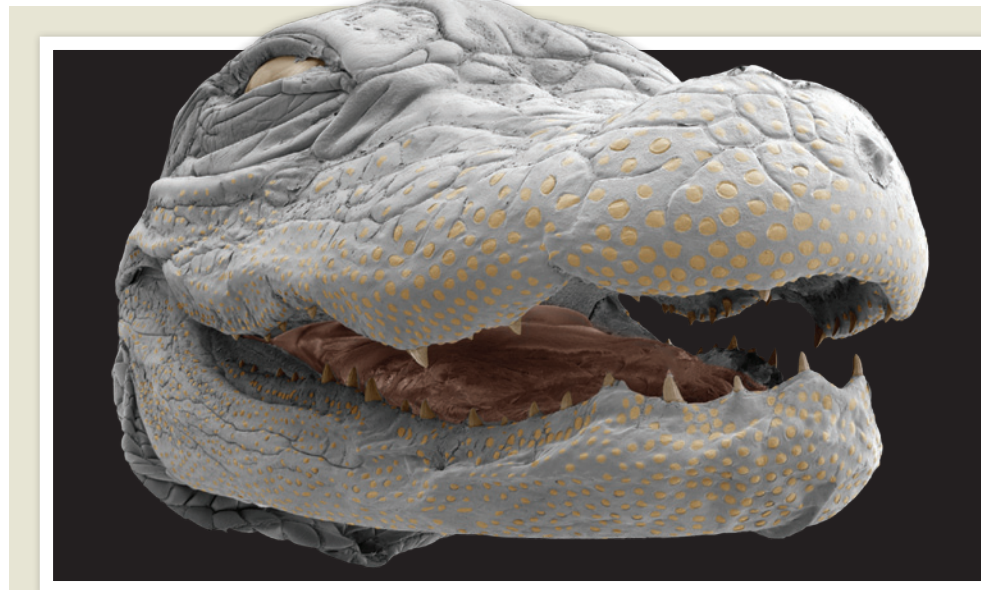
OCEANOGRAPHY

Thinning ice more fragile and mobile

The pronounced thinning of Arctic sea ice has made the ice pack more brittle and susceptible to wind drift.

The volume of Arctic sea ice decreased by one-third during 2007–11 compared with the 1979–2006 mean. In a model simulation, Jinlun Zhang at the University of Washington in Seattle and his colleagues demonstrate that the decline in volume substantially reduces the mechanical strength of the ice, thus boosting ice-drift speed and deformation rates.

Forecasts of ice-edge locations will become more challenging as the thinning



D. LEITCH/VANDERBILT BRAIN INST.

ZOOLOGY

Thick-skinned but sensitive

Crocodiles and alligators may sense their prey using tiny bumps on their mouths that are highly sensitive to touch.

Kenneth Catania and Duncan Leitch at Vanderbilt University in Nashville, Tennessee, investigated the raised bumps — called integumentary sensory organs — in 18 American alligators (*Alligator mississippiensis*; hatchling pictured) and 4 Nile crocodiles (*Crocodylus niloticus*). Confocal microscopy revealed that the bumps (pictured yellow), which are packed

most tightly around the teeth and mouth, share similar structures with tactile skin receptors in mammals. The bumps seemed to be insensitive to electrical current or water salinity, but showed nerve responses when stimulated with a range of levels of force — responding to low levels with a sensitivity exceeding that in primate fingertips.

The authors suggest that the animals use the sensitive bumps to locate prey, and to identify food and other items inside their mouths.

J. Exp. Biol. 215, 4217–4230 (2012)

and weakening of sea ice leads to a state of free drift, the authors note.

Geophys. Res. Lett. <http://dx.doi.org/10.1029/2012GL053545> (2012)

NEUROSCIENCE

Blind reading with sounds

Blind adults taught to ‘read’ using sounds that represent letters use the same area of the brain’s visual cortex that sighted humans use when reading.

Using a program that ‘describes’ images in sound,

Amir Amedi at the Hebrew University of Jerusalem in Israel and his team trained eight congenitally blind people to decipher the shapes of letters and objects such as faces and tools. The authors then imaged participants’ brains as they listened to sounds associated with letters or other objects. When the volunteers read using the sounds, they activated the same part of the visual cortex — the visual word form area (VWFA) — as sighted controls did when viewing letters.

The work shows that the VWFA is not dependent on visual information alone,

and suggests that the visual cortex may be trained to help recognize the environment, even in those who are blind from birth.

Neuron 76, 640–652 (2012)

MATERIALS

Counting loops in gels

Polymer networks, such as rubber and nylon, consist of linked chain-like or branched molecules that almost always contain loops — structural imperfections that weaken a network’s connectivity and

lower the material's elasticity.

Jeremiah Johnson and his colleagues at the Massachusetts Institute of Technology in Cambridge have developed a method to count the number of the most common loops in polymeric materials. The authors broke a hydrogel, a type of polymer network that soaks up water, into quantifiable fragments that reflected the connectivity of the original network, then used mass spectrometry to count the loops. They found that too many loops prevented the gel from forming.

The researchers say they are now using their method to correlate the effects of loops on the mechanical properties of a variety of polymer networks.

Proc. Natl Acad. Sci. USA
<http://dx.doi.org/10.1073/pnas.1213169109> (2012)

ANIMAL BEHAVIOUR

Wrens learn as embryos in the egg

A single song element is all that superb fairy-wren nestlings need to include in their begging calls to get fed by their mothers, and, in an unusual example of prenatal learning, the nestlings seem to learn this 'password' as embryos.

Adult superb fairy-wrens (*Malurus cyaneus*; pictured) use these begging calls to distinguish their offspring from those of two cuckoo species that often invade their nests. Sonia Kleindorfer at Flinders University in Adelaide, Australia, and her team analysed recordings of the fairy-wren calls and found that each nest had a common begging call different from those of all other

nests. That call contained a signature element also found in the call the mother made while incubating her eggs. When the team swapped eggs around across 22 nests, nestlings from those eggs begged using the calls of their foster, not their biological, mothers, suggesting that the calls were learned.

Curr. Biol. <http://dx.doi.org/10.1016/j.cub.2012.09.025> (2012)

NEUROTECHNOLOGY

Brain-machine does the two-step

Brain-machine interfaces (BMIs) detect and use brain activity to perform an intended task, and could be invaluable to people with paralysis. Typically, BMIs are able to process only single movements, but one developed by Ziv Williams at Harvard Medical School in Boston, Massachusetts, and his colleagues can control a series of motions — potentially expanding the complexity of tasks that BMIs can perform.

The team recorded brain activity in monkeys that were trained to move a computer cursor with their paws to each of two areas on a screen in a particular order. This revealed activity in two distinct groups of neurons in the brain's premotor cortex that was associated with each of the upcoming movements. The authors then programmed a computer to decode this signal from the brain and found that the mind-controlled computer moved the cursor at about the same speed that the monkeys achieved with their paws.

Nature Neurosci. <http://dx.doi.org/10.1038/nn.3250> (2012)

NEUROSCIENCE

When neurons mature too early

A genetic mutation linked to intellectual disability and autism causes the premature formation of functional connections between brain cells during a crucial window of development early in life.

COMMUNITY CHOICE

The most viewed papers in science

CHEMISTRY

Mega-MOF's super surface

HIGHLY READ
on pubs.acs.org
in September

Metal-organic frameworks, or MOFs, are of interest for applications such as catalysis and gas storage. Researchers now report a method that allowed them to synthesize these porous crystals with record-breaking surface areas.

Omar Farha at Northwestern University in Evanston, Illinois, and his colleagues created two copper-based MOFs, each with a surface area of approximately 7,000 square metres per gram. To help boost surface area, they used supercritical carbon dioxide to activate the MOFs, avoiding framework collapse, which can occur when the solvents used in MOF synthesis are removed.

Moreover, the authors calculated that by using acetylenes, rather than more bulky phenyls, as links in their framework, they could further increase the theoretical maximum surface area of MOFs to as high as 14,600 square metres per gram, roughly 40% higher than some previous estimates, the team suggests.

J. Am. Chem. Soc. 134, 15016–15021 (2012)

Mutations that inactivate one copy of the gene *SYNGAP1* often cause intellectual disability in humans. Gavin Rumbaugh of the Scripps Research Institute in Jupiter, Florida, and his team found that mice with a similar mutation produce neurons that mature too quickly after birth and become overactive in a brain region important for cognitive function. Mice with one copy of *SYNGAP1* have memory problems and are prone to seizures — a symptom in humans with the mutations.

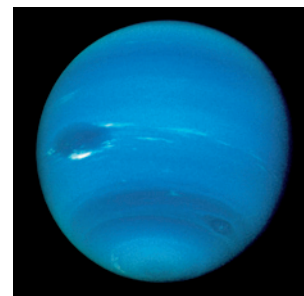
Correcting the mutation in mice after this developmental period had little effect on the symptoms, and introducing the mutation into adult mice did not affect neuronal function — suggesting that the activity of the *SYNGAP1* protein during this developmental window has long-lasting effects.

Cell 151, 1–15 (2012)

ASTRONOMY

More co-orbiters for Neptune

Some astronomers think that Neptune (pictured) can no longer capture objects whose



NASA/SPL

orbits around the Sun are similar to its own. But Carlos and Raúl de la Fuente Marcos at the Complutense University of Madrid in Spain, have discovered that four objects originally classified as minor planets are actually co-orbiters that joined Neptune's orbit as recently as 50,000 years ago.

The work brings to 14 the number of objects that, like Neptune, orbit the Sun every 165 years. The four latest objects are not in the plane of the Solar System and follow eccentric paths. One is likely to diverge from its current path just 2,000 years from now.

Astron. Astrophys. 547, L2 (2012)

NATURE.COM

For the latest research published by Nature visit:

www.nature.com/latestresearch



D. WATTS/NATUREPL.COM

RESEARCH

Stem-cell faker

A scientist who last month fabricated a story about using induced pluripotent stem cells to treat patients with heart failure (see *Nature* **491**, 7–8; 2012) has retracted two of his papers. Hisashi Moriguchi, of the University of Tokyo, retracted research in *Scientific Reports*. He also withdrew his claim to be affiliated with Massachusetts General Hospital and Harvard Medical School, and to have received approval for his work from an institutional review board.

Expensive errors

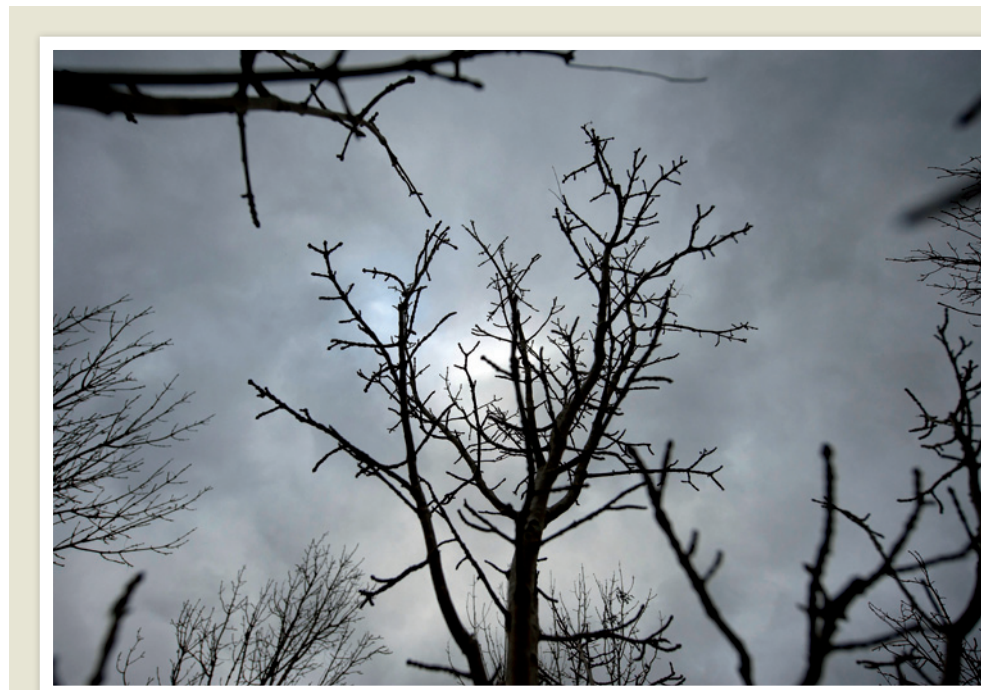
Mistakes in accounting for research projects funded by the European Union in 2011 may have amounted to 3% of the total spending in this area — around €360 million (US\$460 million), according to a report published on 6 November. The European Court of Auditors found that “over-declaration of costs by beneficiaries” was the most common mistake by claimants, but said that most errors were likely to be unintentional. See go.nature.com/nc69zw for more.

Snow survey starts

The Solid Precipitation Intercomparison Experiment, an international two-year project to measure the depth of snow and the amount of snowfall at 15 locations around the world, kicks off this week. See page 312 for more.

Ethics code

Members of the American Anthropological Association overwhelmingly approved new ethical guidelines for research, the professional body announced on 7 November. The rewrite was prompted by a controversy over anthropologists’ participation



B. CLARKE/GETTY

Killer fungus claims UK ash trees

‘Ash dieback’ is likely to kill nearly all of the United Kingdom’s 90 million ash trees, despite the announcement of a plan to control the fungal disease on 9 November. Diseased trees in nurseries will be destroyed, and an import ban

will stay in place. The fungus (*Hymenoscyphus pseudoalbidus*) was initially misidentified by mycologists, making it hard for the European Union to control its spread across the continent. See go.nature.com/anzy6u for more.

in US military efforts in Iraq and Afghanistan (see go.nature.com/so2doa). The new code is rooted in more flexible principles, such as avoiding harm to research subjects and obtaining their informed consent, and not the hard-and-fast rules that the previous ethics code had included.

federal budgets. Scientists hope that the new centre, which will ally the Charité university clinic with the Max Delbrück Center for Molecular Medicine, can rival research powerhouses in the United States and Britain. See page 317 for more.

Open-access funds

Britain’s research-funding agencies will spend more than £100 million (US\$159 million) over the next five years to pay for work that they funded to be free to read, the agencies announced on 8 November. The grants will come out of the United Kingdom’s science budget and will be awarded to universities and other research institutions. Research

Councils UK (an umbrella body for seven agencies that spend a total of £2.8 billion a year) has said that from April 2013, all work funded by the agencies must be published in an open-access format — but the grants for next year are sufficient to cover only 45% of papers. See go.nature.com/d28egt for more.

Fukushima clean-up

The company responsible for decontaminating the ruined Fukushima Daiichi nuclear power plant in Japan says that the cost of doing so will soar. The Fukushima plant suffered meltdowns in three reactors on 11 March 2011 after being stricken by an earthquake and tsunami. On 7 November,

POLICY

Berlin’s biomedicine

Berlin’s mayor Klaus Wowereit announced a deal on 6 November to form the Berlin Institute of Health, which will receive more than €300 million (US\$380 million) in extra funding over the next five years, 90% of it from

LIU JIANSHENG/PHOTOSHOT the Tokyo Electric Power Company, which owns the plant, announced that cleaning up the ruined reactors and surrounding countryside could cost ¥10 trillion (US\$126 billion) — double the size of the clean-up fund set aside by the government.

Space drive

The British contribution to the European Space Agency will rise by £60 million (US\$93 million) per year to £240 million, the UK chancellor George Osborne said in a speech at the Royal Society in London on 9 November. He also named eight technological areas in which the United Kingdom could be a world leader, including satellites and space, synthetic biology, advanced materials and regenerative medicine. See go.nature.com/mcyniw for more.

Fat tax abandoned

Denmark has abolished a tax on high-fat foods, one year into its controversial attempt to make its population healthier. The tax, which added 16 kroner (US\$2.70) for every kilogram of saturated fat in high-fat products, drove up food prices and put jobs at risk, the Danish Ministry of Taxation said. “We have to try improving the public health by other means,” said food minister Mette Gjeskov.

The country has also rejected a proposed tax on high-sugar foods. See go.nature.com/hkgys for more.

Advice at risk

A letter calling for the budget of the UK Parliamentary Office of Science and Technology (POST) to be protected from cuts was published on 7 November, signed by two former British science ministers, a Nobel laureate and many UK learned societies. POST, which provides politicians with analyses of scientific issues, could face cuts of up to £98,000 (US\$156,000), or 17% of its total budget. See go.nature.com/p55ezf for more.

PEOPLE

Palaeontologist dies

Farish Jenkins, a palaeontologist at Harvard University in Cambridge, Massachusetts, died on 11 November, aged 72. Jenkins was a noted science communicator and lecturer. His discovery with colleagues of the fossil remains of *Tiktaalik roseae* — a four-legged, fish-like creature that seemed to show the evolutionary transition of vertebrates from water to land — appeared on the 6 April 2006 cover of *Nature* and sparked a media frenzy. See go.nature.com/y5ehem for more.



Energetic farewell

Hu Jintao, China's outgoing president, said that protecting the environment must be a top priority for the country's next government. In his opening speech to the 18th National Congress in Beijing on 8 November, he told delegates that China must set a ceiling on its energy consumption. Hu (pictured) is likely to be succeeded this week as the head of China's Communist Party by Xi Jinping, who will then become the country's president in March 2013.

BUSINESS

Vaccine tribulations

A large phase III clinical trial of a malaria vaccine candidate, RTS,S/AS01, has reported disappointing results in infants who received their first injection between 6 and 12 weeks of age. The vaccine reduced the number of episodes of clinical malaria over 12 months by around a third for these babies, the key

COMING UP

20–21 NOVEMBER

The European Space Agency council meets in Naples, Italy, to put together a multi-year spending plan. www.esa.int

22–23 NOVEMBER

At a European Council meeting in Brussels, heads of state negotiate the European Union's budget for 2014–20, including the final amount apportioned to research. Around €80 billion (US\$100 billion) has been proposed for the research framework (see *Nature* **489**, 188–189; 2012). go.nature.com/d7rq5u

age group targeted by the trial. The protection is lower than that observed in a previous, smaller phase II trial of babies in the same age group and less than the 55% reported last year in a group of children who were vaccinated at between 5 and 17 months of age. See go.nature.com/gmw9ib for more.

Sequencing stir

Genetic-testing company 23andMe, based in Mountain View, California, will make anonymized customer data available to researchers, it said at the annual meeting of the American Society of Human Genetics in San Francisco, California, last week. At the same event, Pacific Biosciences of Menlo Park, California, announced technology improvements that allow genomic sections at least 5,000 base pairs long to be sequenced in a single read, compared with a few hundred base pairs for other technologies. See go.nature.com/jqfr3k for more.

► **NATURE.COM**

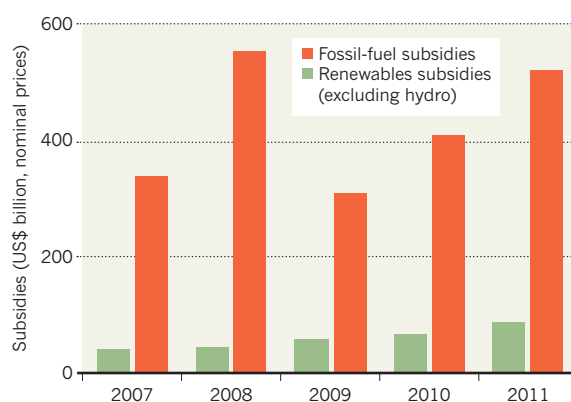
For daily news updates see: www.nature.com/news

TREND WATCH

Subsidies to lower the price of renewable energies, such as solar and wind, rose 24% on 2010 values to reach US\$88 billion in 2011, according to the International Energy Agency's *World Energy Outlook 2012*, released on 12 November. But that support was less than a sixth of the \$523 billion used to artificially reduce the cost of oil, natural gas and coal last year. Fossil-fuel subsidies largely track the price of crude oil. See go.nature.com/i5jt1v for more on this year's energy outlook.

FOSSIL-FUEL SUBSIDIES DWARF RENEWABLES AID

Higher energy prices and rising consumption of oil and gas led to a recent surge of subsidies for fossil fuels to US\$523 billion.



NEWS IN FOCUS

ALTERNATIVE MEDICINE Doubts greet positive outcome of heart therapy **p.313**

GENOMICS Sow's DNA yields silk purse for researchers **p.315**



GERMANY Creative funding for a biomedical powerhouse in Berlin **p.317**

BIOMEDICINE The man who challenged fertility dogma **p.318**



president has “taken major steps to reduce our carbon consumption, including setting higher fuel-efficiency standards for cars and trucks”. When he accepted that endorsement, Obama acknowledged that “climate change is a threat to our children’s future, and we owe it to them to do something about it”.

Yet this new opportunity to confront climate change and invest in science and technology comes with towering obstacles. The election did not end the polarization of Congress — Republicans retained control of the House of Representatives, and the Democrats only slightly strengthened their narrow majority in the Senate. And the ‘fiscal cliff’ looms large — automatic tax increases and spending cuts, the legacy of earlier budget battles, will hit on 2 January unless the outgoing Congress finds a way to avert them in the session that begins this week (see *Nature* **487**, 414–415; 2012).

The cuts, totalling some US\$136 billion, would apply to all discretionary spending next year, including defence, and would eat deeply into federal science budgets (see ‘At the precipice’). Congressional leaders expect Obama to play an active part in brokering a deal to avoid the fiscal cliff, which economists say could plunge the fragile US economy back into recession. The outcome will foreshadow Obama’s prospects for achieving other objectives — including those relevant to science — during his second term.

CLIMATE OPPORTUNITY

Obama may have to develop his climate plans without some high-profile lieutenants. Energy secretary Steven Chu and Environmental Protection Agency (EPA) administrator Lisa Jackson are both rumoured to be stepping down. During Obama’s first term, both became lightning rods for Republican attacks — Chu for his role in approving a \$535-million government loan guarantee to Solyndra, a solar-energy company that later went bankrupt, and Jackson for implementing greenhouse-gas regulations. But even without Chu or Jackson, the administration’s approach to renewable energy and global warming would change very little, says Michael Gerrard, director of the Center for Climate Change Law at Columbia University in New York. “I suspect we would have continuity in the broad policy approaches.”

Obama’s election victory, combined with growing alarm in the United States over ►

Re-elected US President Barack Obama has won four more years in which to cement his legacy.

POLICY

Obama reasserts research focus

But ‘fiscal cliff’ threatens science and climate goals.

BY ERIC HAND, IVAN SEMENIUK, JEFF TOLLEFSON AND MEREDITH WADMAN

Speaking last week in Washington DC, US President Barack Obama reminded voters of the plan they had effectively endorsed by re-electing him. One of his key objectives, Obama said, would be to ensure that the United States “is a global leader in research and technology and clean energy, which will attract new

companies and high-wage jobs to America”.

A different objective will be in the spotlight this week when Obama visits New York, a city still recovering from the damage caused by Hurricane Sandy on 29 October. Climate change could make storms like Sandy more common in the future. And New York’s mayor, Michael Bloomberg, cited climate concerns when he endorsed Obama for re-election (see *Nature* **491**, 167–168; 2012), saying that the

PEOPLE'S CHOICE

Science at stake in state proposals

During last week's national election, US voters also weighed in on state ballot measures that affect research.

Hope for higher education

California voters approved temporary tax increases expected to generate US\$6 billion annually for schools and community colleges through 2017, with smaller revenues through 2019. Passage of the measure also halted imminent state budget cuts that would have cost the University of California (UC) and the California State University system \$250 million each over one year. UC administrators expressed optimism that the measure heralds a more favourable fiscal and political climate for the institution, which has weathered years of state funding cutbacks.

"Education and research are multi-year endeavours," says Keith Yamamoto, vice-chancellor for research at UC San Francisco. "The impact of a lack of stable funding is substantial."

Steve Montiel, a UC spokesman, called the measure "a significant step towards the prospect of financial stability".

Experimentation with marijuana

Colorado and Washington became the first states to legalize marijuana for non-medical purposes, permitting purchase and use of the drug by adults aged 21 and older. Residents of Colorado can also possess up to six marijuana plants.

The new measures set the stage for a legal battle: federal law prohibits the substance, and the US Drug Enforcement Administration says that its policy "remains unchanged". Opponents fear that looser marijuana laws could lead to more drug abuse in the United States. In studies of

drug use in the Netherlands since the country's de facto legalization of marijuana in 1976, Robert MacCoun, a drug-policy expert at the University of California, Berkeley, has found only modest increases in marijuana use and no significant escalation to harder drugs (R. MacCoun and P. Reuter, *Science* **278**, 47–52; 1997).

"Of course, the Netherlands is a different country", but it provides some of the only available data worldwide, MacCoun says. The Washington measure would direct some marijuana tax revenue towards drug-abuse research.

No labelling for genetically modified food

California voters rejected a measure that would have made theirs the first state to require labelling of foods containing genetically modified organisms (GMOs).

Bob Goldberg, a plant geneticist at UC Los Angeles, who co-authored an argument against the proposal in the state's voting guide, calls the measure "ideological and not evidence-based".

Marion Nestle, a food-politics expert at New York University, argues that labelling might have promoted greater consumer trust in GMOs. "Not having a choice induces paranoia in people," she says.

A poll conducted by the California Business Roundtable, a business-advocacy group based in Sacramento, and the School of Public Policy at Pepperdine University in Malibu, California, found that public support for labelling was at around 65% in August, then plummeted in the month before the election. The opposing campaign — backed heavily by Monsanto, a producer of genetically modified seeds based in St Louis, Missouri — escalated its television advertising in the final weeks. **Helen Shen**

starting point, but we think it could become extremely attractive if the lawmakers begin to run out of options for generating revenue," says Mark Muro, a senior fellow at the Brookings Institution, a non-partisan think tank in Washington DC.

Although the idea of any tax — let alone one on carbon — is anathema to most conservatives on Capitol Hill, the idea offers something for everybody in the current budget crunch. A tax would reduce emissions and raise revenue for energy research and development — outcomes that environmentalists would welcome — and it would generate extra revenue that conservative lawmakers could use to offset lower tax rates on individuals and businesses.

In August, economists at the Massachusetts Institute of Technology (MIT) in Cambridge reported that the United States could raise \$1.5 trillion over ten years and reduce emissions to 14% below 2006 levels by 2020 by instituting a tax of \$20 per tonne of carbon in 2013 and increasing it by 4% a year. Less than 20% of that revenue would be enough to fund a massive boost in federal investments in clean-energy research and development — from \$3.8 billion in 2012 to \$30 billion annually, Muro says. The administration has yet to weigh in on the idea, and some on Capitol Hill think that is just as well. If pushed prematurely, the idea could wither in the political spotlight before lawmakers have a chance to fully consider its merits. "We're trying to generate interest," a senior House aide told *Nature*. "The more discussion there is, the better."

FUNDING FEARS

Basic research has historically fared well under Democratic and Republican administrations, but many university administrators were nonetheless relieved by the election's outcome. They wondered whether Republican candidate Mitt Romney would side with other Republicans — including Paul Ryan, his running mate and chairman of the House budget committee — who advocate severe spending cuts to government programmes as a way of reducing the deficit. "Thank God we don't have to find out," says Stewart Smith, dean for research at Princeton University in New Jersey.

Universities that receive grant money from federal science agencies are nonetheless bracing themselves for the fiscal cliff that awaits if Congress cannot reach a budget deal or find a way to extend the bargaining window

before 2 January. The administration's Office of Management and Budget estimates that most funding agencies would have their budgets slashed by 8.2% in the absence of an agreement. Claude Canizares, vice-president for research at MIT, says that the result

▶ severe weather events, such as Hurricane Sandy and the severe drought in the Midwest this summer, could bolster efforts to curb carbon emissions. Jackson laid the foundation for such reductions after the US Supreme Court ruled in 2007 that carbon dioxide is a pollutant, which allowed the EPA to regulate the gas under the Clean Air Act. Jackson went on to craft the first US greenhouse-gas regulations for vehicles, and in March proposed a rule that would effectively ban the construction of new coal-fired power plants unless they are equipped to capture and sequester roughly 50% of the carbon dioxide they emit. The agency next plans to propose rules for existing power plants, then oil refineries. The details of those rules are unclear. The EPA could, for

example, set energy-efficiency standards for different types of power plant or take a more flexible approach that would let states — which normally implement air-quality rules — decide how to proceed.

There is also an outside possibility that Congress's struggle to avoid the fiscal cliff could bring another approach to the fore: a carbon tax. To avoid the automatic spending cuts and tax rises, lawmakers must find new ways to raise revenue. So far, the House has focused on closing tax loopholes, but those efforts are expected to come up short. As an alternative, a carbon tax has been quietly gaining traction in policy meetings, even among some conservatives.

"A carbon tax is certainly not likely as a



NATURE.COM
Visit *Nature's*
election special:
[www.nature.com/
election2012](http://www.nature.com/election2012)

LEFT: T. WILLIAMS/CO ROLL CALL/GETTY
RIGHT: S. SHAVER/UP/EYEVINE

US Environmental Protection Agency administrator Lisa Jackson (left) and energy secretary Steven Chu (right) drew much Republican ire in 2008–12.

would be a loss of \$40 million in research revenue at the university, which drew \$619 million in research support in 2011. An 8% cut in funding to the US Department of Energy (DOE) would mean an 8% loss in staff at the DOE Princeton Plasma Physics Laboratory, of which Smith will become vice-president in January.

Advocates for biomedical research are equally concerned about the prospect of the cuts. “I don’t know how you spare anyone or anything” at the US National Institutes of Health (NIH), says Jennifer Zeitzer, director of legislative affairs at the Federation of American Societies for Experimental Biology in Bethesda, Maryland. The automatic cuts would slash the agency’s \$30.7-billion budget by \$2.5 billion, a portion of which would be exacted from every NIH institute and centre.

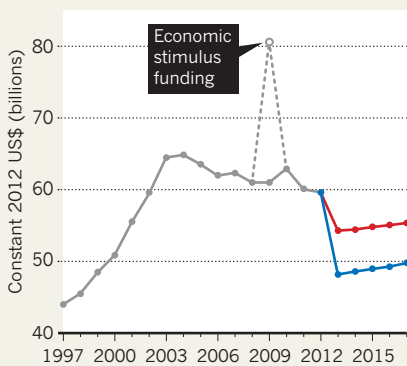
That prospect, in combination with other pressures on medical research budgets, is “chilling”, says Ann Bonham, chief scientific officer at the Association of American Medical Colleges in Washington DC. She notes that the Patient Protection and Affordable Care Act, the health-care-reform law that is the signature policy achievement of Obama’s first term and will ultimately extend health insurance to more than 30 million now-uninsured US citizens, also mandates a \$155-billion cut in government payments to hospitals. That could hurt the large teaching hospitals that support much US medical research (see *Nature* 487, 13–14; 2012). Bonham worries that the confluence of stresses will consign the biomedical research enterprise to “death by 1,000 cuts”.

Supporters of Obama’s health-care reform argue that it will eventually curb soaring US

government outlays for health care, thereby shrinking the deficit and producing more revenue for agencies, including the NIH. If the plan works, the reform is “likely to be a long-term gain for research”, says Ezekiel Emanuel, a medical ethicist and health-policy expert at the University of Pennsylvania in Philadelphia, who was the senior health-policy adviser to the White House budget office from 2009–11.

AT THE PRECIPICE

Annual US funding for non-military research and development will drop sharply if government-wide budget cuts take effect (red). If defence spending is spared, even deeper cuts will apply (blue).



Science could also benefit from another Obama goal: immigration reform. One result of such reform could be more ‘H-1B’ visas for foreign scientists and engineers, a need that both candidates emphasized during the campaign.

In the meantime, Obama’s science agenda

will depend on cooperation with a Congress that includes some new faces. Last week’s general election not only returned him to office and decided state ballot measures (see ‘People’s choice’); it also marked the end of some key lawmakers’ terms. The chairmanship of the Senate energy and water committee is being vacated by the retiring Jeff Bingaman, a Democrat from New Mexico. Some observers expect Ron Wyden, a Democrat from environmentally friendly Oregon, to fill the spot. Although generally liberal, Wyden is known for crossing the aisle to work with Republicans. “Wyden is a very thoughtful guy; he likes to think outside the box,” says Michael Lubell, director of public affairs for the American Physical Society in Washington DC.

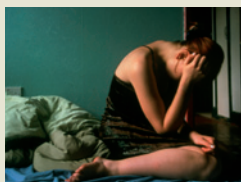
In the House, it’s unclear who will chair the spending subcommittee that funds the NIH. The current chairman, Denny Rehberg of Montana, relinquished his House seat in a failed bid for the Senate. The science, space and technology committee is losing its chairman, Ralph Hall of Texas, who is stepping down in accordance with Republican term limits. Jim Sensenbrenner of Wisconsin, Dana Rohrabacher of California and Lamar Smith of Texas all want the job. Lubell says that his money is on Smith, a co-sponsor of patent-reform legislation who has also tried to make it easier for foreign graduates with science degrees to remain in the United States.

For now, Obama’s victory has created an opening for compromise after two years of Congressional gridlock. With the fiscal clock ticking, the coming weeks may well set the tone for the next four years. ■ SEE EDITORIAL P. 301

SOURCE: AAAS/CONGRESSIONAL BUDGET OFFICE


**MORE
ONLINE**

TOP STORY



Infant stress affects teen brain
go.nature.com/a28sgk

MORE NEWS

- Corals under attack summon support from friendly fish go.nature.com/16illy
- Wrens teach their eggs to sing go.nature.com/4cmkih
- Drought hastened Maya decline go.nature.com/lfbhpt

ON THE BLOG



Pre-Columbian fossil collectors unearthed
go.nature.com/fujihy

J. VARNEY/SPL

L. W. FALVEY & B. MCCLARIN



Kasprowy Wierch, a summit in Poland, will soon get new snow-measuring instruments.

CLIMATE SCIENCE

Snow survey hopes for avalanche of data

More accurate snowfall measurements could improve climate models and estimates of water resources.

BY JANE QIU

Mountains are barometers of climate change, but some of the simplest questions about them are the hardest to answer. How much snow coats their peaks and slopes, for example? And how do these frosty shrouds alter from year to year? This week, an international programme kicks off to answer those questions.

In a two-year project called the Solid Precipitation Intercomparison Experiment (SPICE), spearheaded by the World Meteorological Organization (WMO), climate scientists will deploy a suite of state-of-the-art snow gauges at 15 sites in geographically and climatically diverse countries around the world, up

to 4,318 metres above sea level (see 'White noise'). The goal is to make accurate measurements of snow depth and snowfall — the most fragile form of precipitation, which can elude or clog simple collecting devices — and come up with recommendations for the best ways to do snow surveys in different parts of the world. The results could improve climate models and help to predict permafrost stability, ecosystem changes and the availability of water resources in the coming decades.

"Snowfall is an important part of the global hydrological cycle," says Roger Atkinson, acting head of the WMO's Instruments and Methods of Observation Programme in Geneva, Switzerland. "If we can't accurately measure the amount of snowfall, then we won't

be able to know how much water we have and how it will change in the future."

Snowfall also "partly determines whether a glacier grows or retreats," says Zhang Yinsheng, a climate scientist at the Chinese Academy of Sciences' Institute of Tibetan Plateau Research in Beijing, who is not involved in the SPICE project. "People have been debating the fate of Himalayan glaciers for a long time, but we don't have a proper grasp of even the basics."

Although researchers can accurately assess some climate parameters such as temperature, pressure, wind speed and humidity, measuring snowfall remains challenging. Snowflakes are light and drift with the wind, and weather conditions can dramatically affect what proportion of snowfall is actually captured by researchers' cylindrical metal gauges, says Rodica Nitu, a meteorological instrument expert at Environment Canada in Toronto who leads the project. And when the temperature is around freezing, the damp snow tends to stick to the rim of the container, soon forming a cap and preventing further collection.

"Undercatch is a serious problem," says Roy Rasmussen, a climate scientist at the US National Center for Atmospheric Research in Boulder, Colorado. This is particularly the case with automatic gauges, which can capture as little as 20% of the actual snowfall, he says.

Unreliable snow readings introduce one of the greatest unknowns to climate models, hampering the ability to predict future changes in water resources and mountain hazards, says Rasmussen. And snowfall, like other forms of precipitation, is expected to increase as the globe warms. Better snow data could help modellers to predict the increase in snowfall, and whether it will be sufficient to offset the increased melting of glaciers.

The last major international effort to measure snow was more than 20 years ago, and "there has been a lot of progress since then," says Rasmussen. One of the main aims of the project is to test the range of recently developed sensors, gauges and windshields. For example, field observations show that shields to reduce the horizontal wind speed above the gauge increase collection enormously. "It's the most important factor for accurate snow measurements," says Rasmussen.

New ways of heating the measuring gauges should also prevent snow capping without causing evaporation or air turbulence that blows the snow away. The field is also switching from manual to automated instruments, enabling continuous measurements over large, hard-to-access areas. Relating the two data sets will make records of snow measurements continuous over time, says Nitu.

Zhang says that the project is timely and important, but that it misses crucial regions such as the Himalayas, where SPICE doesn't have a testing site. Early next year, Zhang and his colleagues will set up a network of

A. RADOSAVLJEVIC/ALAMY

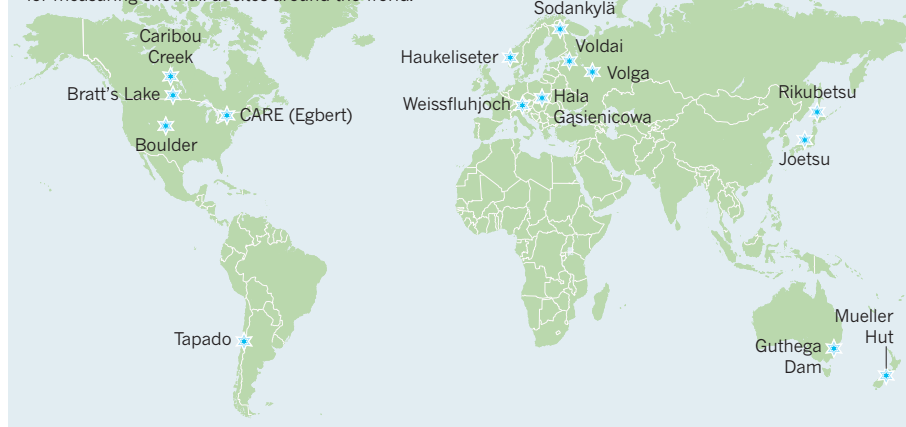
stations across the Tibetan plateau and surrounding mountain ranges, at altitudes up to 6,000 metres, which will gather accurate snow measurements across the region that could augment SPICE's results.

In the longer term, however, "there will never be enough ground measurements to cover an entire mountain", says Michael Lehning, a climate scientist at the Swiss Federal Institute for Snow and Avalanche Research in Davos-Dorf, who is involved in the project. Results from SPICE will be used to calibrate airborne and satellite-based sensors, which use techniques such as microwave, radar and laser ranging to survey much larger areas.

"The idea is to push remote sensing to be accurate enough for use in mountains," says Lehning. "It's still a long way off, but SPICE is a good starting point." ■

WHITE NOISE

The SPICE project aims to test the most advanced techniques for measuring snowfall at sites around the world.



ALTERNATIVE MEDICINE

Chelation-therapy heart trial draws fire

Critics not persuaded that metal-snaring treatment works.

BY EWEN CALLAWAY

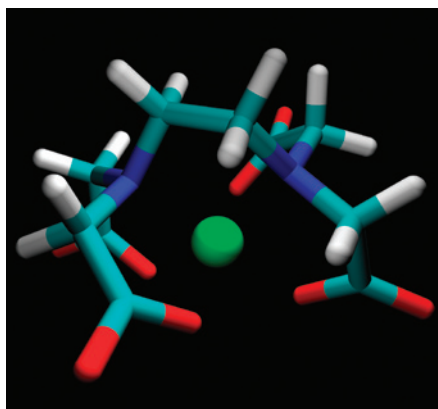
With millions of Americans regularly using complementary medicines, researchers usually applaud efforts to test and debunk folk treatments such as echinacea, a herbal supplement often deployed against the common cold. But what if a trial shows that an alternative therapy might work?

That is the case for a study funded by the US National Center for Complementary and Alternative Medicine (NCCAM), part of the National Institutes of Health (NIH) in Bethesda, Maryland. The trial hints that a fringe therapy intended to sop up metal ions in the blood might reduce participants' risk of heart attack. Critics are attacking both the rigour of the study and the records of some of its investigators, complicating the NCCAM's efforts to answer charges from some researchers that it funds quackery, and raising questions about whether the centre's US\$128-million annual budget is being spent wisely.

The Trial to Assess Chelation Therapy (TACT) was a 10-year, \$31.6-million study involving 1,708 participants at 134 centres. It aimed to test whether weekly infusions of a salt of ethylenediaminetetraacetic acid (EDTA) can lower the risk of repeat heart attacks. EDTA is a chelating agent: the molecule is peppered with electron-rich nitrogen and oxygen atoms, which can grab and hold onto positive metal ions (see picture). The US Food and Drug

Administration has approved one salt, calcium disodium EDTA, to treat lead poisoning.

Proponents of chelation therapy for heart disease initially speculated that EDTA could also cleanse the blood of calcium ions, a component of the atherosclerotic plaques that block blood vessels. But evidence against that hypothesis led them to suggest alternative mechanisms, for example that the molecule captures other metals, preventing heart-damaging inflammation. In spite of the uncertainty, the treatment is already big business: a 2007 US government survey estimated that, every year, 110,000 Americans undergo chelation therapy,



The electron-rich oxygen (red) and nitrogen (dark blue) atoms in ethylenediaminetetraacetic acid can grab and hold onto positive metal ions (green).

which can cost thousands of dollars per course.

According to TACT, which the NCCAM co-funded along with the National Heart, Lung, and Blood Institute (NHLBI), the therapy shows signs of working. On 4 November at the annual meeting of the American Heart Association in Los Angeles, California, trial leaders reported that 26% of patients who received infusions of disodium EDTA went on to suffer a heart attack, stroke or other heart problem, compared with 30% of patients on a placebo — a statistically significant difference.

Many medical researchers were quick to question the results. Perplexingly, the benefit was observed only among participants with diabetes, and 30% of participants dropped out of the trial, undermining comparison between the treatment and placebo. Critics also note that nearly two dozen trial co-investigators have been disciplined by state medical boards for infractions ranging from insurance fraud to providing ineffective treatments. "They offer aromatherapy, crystal therapy and every imaginable wacky form of medicine. You can't do high-quality research at sites like that," says Steven Nissen, a cardiologist at the Cleveland Clinic in Ohio. "We wasted \$30 million and 10 years on an unreliable study."

He worries that the research will be used to support unapproved chelation therapies, which have been linked to heart attacks and death. "Public harm is going to come out of this. People are going to get bilked out of a lot of money. People are going to die."

Kimball Atwood, an anaesthesiologist at Tufts University School of Medicine in Boston, Massachusetts, and one of TACT's most vociferous critics, argues that the trial has been troubled from the beginning. In a paper titled 'Why the NIH Trial to Assess Chelation Therapy (TACT) should be abandoned' (K. C. Atwood *et al. Medscape J. Med.* **10**, 115; 2008), he claimed that trial proponents had mischaracterized earlier studies of chelation therapy ►

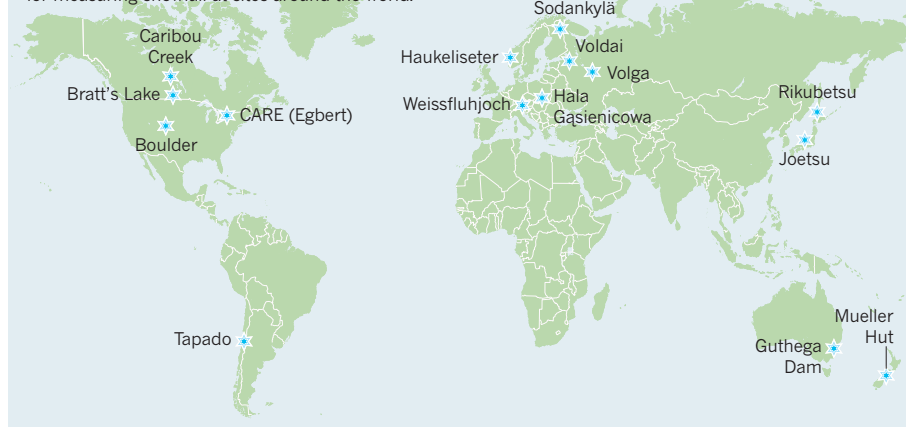
stations across the Tibetan plateau and surrounding mountain ranges, at altitudes up to 6,000 metres, which will gather accurate snow measurements across the region that could augment SPICE's results.

In the longer term, however, "there will never be enough ground measurements to cover an entire mountain", says Michael Lehning, a climate scientist at the Swiss Federal Institute for Snow and Avalanche Research in Davos-Dorf, who is involved in the project. Results from SPICE will be used to calibrate airborne and satellite-based sensors, which use techniques such as microwave, radar and laser ranging to survey much larger areas.

"The idea is to push remote sensing to be accurate enough for use in mountains," says Lehning. "It's still a long way off, but SPICE is a good starting point." ■

WHITE NOISE

The SPICE project aims to test the most advanced techniques for measuring snowfall at sites around the world.



ALTERNATIVE MEDICINE

Chelation-therapy heart trial draws fire

Critics not persuaded that metal-snaring treatment works.

BY EWEN CALLAWAY

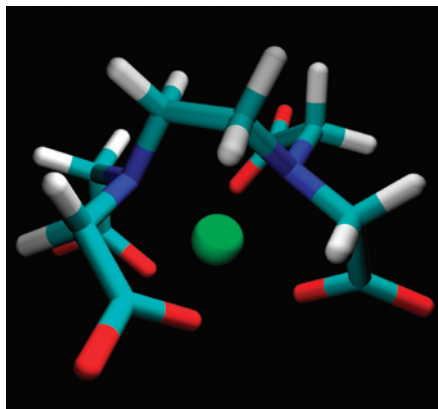
With millions of Americans regularly using complementary medicines, researchers usually applaud efforts to test and debunk folk treatments such as echinacea, a herbal supplement often deployed against the common cold. But what if a trial shows that an alternative therapy might work?

That is the case for a study funded by the US National Center for Complementary and Alternative Medicine (NCCAM), part of the National Institutes of Health (NIH) in Bethesda, Maryland. The trial hints that a fringe therapy intended to sop up metal ions in the blood might reduce participants' risk of heart attack. Critics are attacking both the rigour of the study and the records of some of its investigators, complicating the NCCAM's efforts to answer charges from some researchers that it funds quackery, and raising questions about whether the centre's US\$128-million annual budget is being spent wisely.

The Trial to Assess Chelation Therapy (TACT) was a 10-year, \$31.6-million study involving 1,708 participants at 134 centres. It aimed to test whether weekly infusions of a salt of ethylenediaminetetraacetic acid (EDTA) can lower the risk of repeat heart attacks. EDTA is a chelating agent: the molecule is peppered with electron-rich nitrogen and oxygen atoms, which can grab and hold onto positive metal ions (see picture). The US Food and Drug

Administration has approved one salt, calcium disodium EDTA, to treat lead poisoning.

Proponents of chelation therapy for heart disease initially speculated that EDTA could also cleanse the blood of calcium ions, a component of the atherosclerotic plaques that block blood vessels. But evidence against that hypothesis led them to suggest alternative mechanisms, for example that the molecule captures other metals, preventing heart-damaging inflammation. In spite of the uncertainty, the treatment is already big business: a 2007 US government survey estimated that, every year, 110,000 Americans undergo chelation therapy,



The electron-rich oxygen (red) and nitrogen (dark blue) atoms in ethylenediaminetetraacetic acid can grab and hold onto positive metal ions (green).

which can cost thousands of dollars per course.

According to TACT, which the NCCAM co-funded along with the National Heart, Lung, and Blood Institute (NHLBI), the therapy shows signs of working. On 4 November at the annual meeting of the American Heart Association in Los Angeles, California, trial leaders reported that 26% of patients who received infusions of disodium EDTA went on to suffer a heart attack, stroke or other heart problem, compared with 30% of patients on a placebo — a statistically significant difference.

Many medical researchers were quick to question the results. Perplexingly, the benefit was observed only among participants with diabetes, and 30% of participants dropped out of the trial, undermining comparison between the treatment and placebo. Critics also note that nearly two dozen trial co-investigators have been disciplined by state medical boards for infractions ranging from insurance fraud to providing ineffective treatments. "They offer aromatherapy, crystal therapy and every imaginable wacky form of medicine. You can't do high-quality research at sites like that," says Steven Nissen, a cardiologist at the Cleveland Clinic in Ohio. "We wasted \$30 million and 10 years on an unreliable study."

He worries that the research will be used to support unapproved chelation therapies, which have been linked to heart attacks and death. "Public harm is going to come out of this. People are going to get bilked out of a lot of money. People are going to die."

Kimball Atwood, an anaesthesiologist at Tufts University School of Medicine in Boston, Massachusetts, and one of TACT's most vociferous critics, argues that the trial has been troubled from the beginning. In a paper titled 'Why the NIH Trial to Assess Chelation Therapy (TACT) should be abandoned' (K. C. Atwood *et al. Medscape J. Med.* **10**, 115; 2008), he claimed that trial proponents had mischaracterized earlier studies of chelation therapy ►

► by suggesting that their results were equivocal and necessitated a larger follow-up. Atwood says that the earlier studies in fact found that the treatment was ineffective at preventing heart attacks.

In 2008, TACT was suspended after regulators learned that subjects were not being given calcium disodium EDTA, as implied on informed-consent forms — instead, they were being infused with the slightly different salt disodium EDTA, for which the FDA had revoked approval. The trial resumed after consent forms were reworded to include warnings, such as “death is a rare complication of EDTA infusions”.

Josephine Briggs, director of the NCCAM, declined to comment on TACT until the results are published in a journal. The principal investigator, cardiologist Gervasio Lamas of Mount Sinai Medical Center in Miami Beach, Florida, says that the study's findings were a surprise and deserve following up. He adds that the trial consent forms were approved by the NIH and multiple institutional review boards. Gary Gibbons, director of the NHLBI, says that his institute stands by the study's methodology.

But critics charge that TACT is simply the latest example of dubious research into unproven therapies supported by the NCCAM. Some argue that even high-quality studies would have little value, because negative results are unlikely to sway ardent practitioners. “Show me one alternative medication or procedure that was studied, found to not work, and was abandoned by practitioners. I’m not aware of any,” says Steven Novella, a neurologist at Yale University in New Haven, Connecticut. Briggs, who previously led the NIH’s kidney-disease research, points out that echinacea sales fell after an NCCAM-funded study showed it was ineffective against colds (R. B. Turner *et al.* *N. Engl. J. Med.* 353, 341–348; 2005). With the centre’s research showing that Americans spend about \$34 billion on alternative medicine each year, “we think it’s really important to bring some science into this,” she says.

Briggs adds that the NCCAM’s critics often misrepresent the centre’s research, focusing on studies of herbal supplements such as lavender oil but ignoring multi-million-dollar grants for more-mainstream science. Among the largest studies funded by the centre this year are a computational analysis of the human microbiome and an effort to use brain imaging to understand and treat chronic back pain.

Novella and other NCCAM critics do praise Briggs for bringing increased accountability to the centre, and for boosting the rigour of the research it funds. But “even if you did pristine research under the NCCAM,” says Novella, “it’s what you’re studying that is the problem.” ■

BIOTECHNOLOGY

Pig geneticists go the whole hog

Genome will benefit farmers and medical researchers.

BY ALISON ABBOTT

T. J. Tabasco is something of a porcine goddess at the University of Illinois, Urbana-Champaign, where her ruddy, taxidermied head looks down from the office wall of geneticist Lawrence Schook. Now she has been immortalized in this week’s *Nature*¹ — not by name, but by the letters of her DNA.

Scientists are salivating. For the past couple of decades they have been slowly teasing information from the pig genome, applying it to breed healthier and meatier pigs, and to try to create more faithful models of human disease. This week’s draft sequence of T. J.’s genome (see page 393), with its detailed annotation — a ‘reference genome’ — will speed progress on both fronts, and perhaps even allow pigs to be engineered to provide organs for transplant into human patients. “Agriculture in particular will benefit fast,” says Alan Archibald of the Roslin Institute in Edinburgh, UK, one of the paper’s lead authors. “The pig industry has an excellent track record for rapid adoption of new technologies and knowledge.”

T. J., a domestic Duroc pig (*Sus scrofa domestica*), was born in Illinois in 2001. The next year, Schook and his colleagues generated a fibroblast cell line from a small piece of skin from her ear and commissioned clones to be created from it, so that they could work on animals all with the same genome. One set of clones was created at the National Swine Resource and Research Center (NSRRC) in Columbia, Missouri, along with genetically engineered pigs with genes added or deleted to mimic human diseases. “Making such pigs has got increasingly easier as knowledge of the genome increases,” says physiologist Randall Prather, a co-director of the NSRRC, which is funded by the National Institutes of Health (NIH).

The NIH launched the NSRRC in 2003 to encourage research in pig disease models. Pigs are more expensive to keep than rodents, and they reproduce more slowly. But the similarities between pig and human anatomy and physiology can trump the drawbacks. For example, their eyes are a similar size, with photoreceptors similarly distributed in the retina. So the pig became the first model for retinitis pigmentosa, a cause of blindness. And four years ago, researchers created a pig model of cystic fibrosis² that, unlike mouse models, developed



T. J. Tabasco, star of the show.

symptoms resembling those in humans.

Geneticist and veterinarian Eckhard Wolf at the Ludwig-Maximilian University in Munich, Germany, has exploited the similarity between the human and pig gastrointestinal system and metabolism — like us, pigs will eat almost anything and then suffer for it — to develop models of diabetes. One pig model carries a mutant transgene that limits the effectiveness of incretin, a hormone required for normal insulin secretion³. Mice with the transgene developed unexpectedly severe diabetes, but the pigs have a more subtle pre-diabetic condition that better models the human disease. “This shows the importance of using an animal with a relevant physiology,” says Wolf.

Pig models are now being developed for other common conditions, including Alzheimer’s disease, cancer and muscular dystrophy. This work will be enriched by the discovery, reported in the genome paper, of 112 gene variants that might be involved in human diseases. Knowledge of the genome is also allowing scientists to try to engineer pigs that could be the source of organs, including heart and liver, for human patients. Pig organs are roughly the right size, and researchers hope to create transgenic pigs carrying genes that deceive the immune system of recipients into not rejecting the transplants.

Back on the farm, early knowledge about the pig genome led to the discovery in 1991 of a gene involved in porcine stress syndrome, in which the stress of overheating, being moved or even having sex causes the animals to die suddenly⁴. It then became possible to test for the ►

► by suggesting that their results were equivocal and necessitated a larger follow-up. Atwood says that the earlier studies in fact found that the treatment was ineffective at preventing heart attacks.

In 2008, TACT was suspended after regulators learned that subjects were not being given calcium disodium EDTA, as implied on informed-consent forms — instead, they were being infused with the slightly different salt disodium EDTA, for which the FDA had revoked approval. The trial resumed after consent forms were reworded to include warnings, such as “death is a rare complication of EDTA infusions”.

Josephine Briggs, director of the NCCAM, declined to comment on TACT until the results are published in a journal. The principal investigator, cardiologist Gervasio Lamas of Mount Sinai Medical Center in Miami Beach, Florida, says that the study's findings were a surprise and deserve following up. He adds that the trial consent forms were approved by the NIH and multiple institutional review boards. Gary Gibbons, director of the NHLBI, says that his institute stands by the study's methodology.

But critics charge that TACT is simply the latest example of dubious research into unproven therapies supported by the NCCAM. Some argue that even high-quality studies would have little value, because negative results are unlikely to sway ardent practitioners. “Show me one alternative medication or procedure that was studied, found to not work, and was abandoned by practitioners. I'm not aware of any,” says Steven Novella, a neurologist at Yale University in New Haven, Connecticut. Briggs, who previously led the NIH's kidney-disease research, points out that echinacea sales fell after an NCCAM-funded study showed it was ineffective against colds (R. B. Turner *et al.* *N. Engl. J. Med.* 353, 341–348; 2005). With the centre's research showing that Americans spend about \$34 billion on alternative medicine each year, “we think it's really important to bring some science into this,” she says.

Briggs adds that the NCCAM's critics often misrepresent the centre's research, focusing on studies of herbal supplements such as lavender oil but ignoring multi-million-dollar grants for more-mainstream science. Among the largest studies funded by the centre this year are a computational analysis of the human microbiome and an effort to use brain imaging to understand and treat chronic back pain.

Novella and other NCCAM critics do praise Briggs for bringing increased accountability to the centre, and for boosting the rigour of the research it funds. But “even if you did pristine research under the NCCAM,” says Novella, “it's what you're studying that is the problem.” ■

BIOTECHNOLOGY

Pig geneticists go the whole hog

Genome will benefit farmers and medical researchers.

BY ALISON ABBOTT

T. J. Tabasco is something of a porcine goddess at the University of Illinois, Urbana-Champaign, where her ruddy, taxidermied head looks down from the office wall of geneticist Lawrence Schook. Now she has been immortalized in this week's *Nature*¹ — not by name, but by the letters of her DNA.

Scientists are salivating. For the past couple of decades they have been slowly teasing information from the pig genome, applying it to breed healthier and meatier pigs, and to try to create more faithful models of human disease. This week's draft sequence of T. J.'s genome (see page 393), with its detailed annotation — a ‘reference genome’ — will speed progress on both fronts, and perhaps even allow pigs to be engineered to provide organs for transplant into human patients. “Agriculture in particular will benefit fast,” says Alan Archibald of the Roslin Institute in Edinburgh, UK, one of the paper's lead authors. “The pig industry has an excellent track record for rapid adoption of new technologies and knowledge.”

T. J., a domestic Duroc pig (*Sus scrofa domestica*), was born in Illinois in 2001. The next year, Schook and his colleagues generated a fibroblast cell line from a small piece of skin from her ear and commissioned clones to be created from it, so that they could work on animals all with the same genome. One set of clones was created at the National Swine Resource and Research Center (NSRRC) in Columbia, Missouri, along with genetically engineered pigs with genes added or deleted to mimic human diseases. “Making such pigs has got increasingly easier as knowledge of the genome increases,” says physiologist Randall Prather, a co-director of the NSRRC, which is funded by the National Institutes of Health (NIH).

The NIH launched the NSRRC in 2003 to encourage research in pig disease models. Pigs are more expensive to keep than rodents, and they reproduce more slowly. But the similarities between pig and human anatomy and physiology can trump the drawbacks. For example, their eyes are a similar size, with photoreceptors similarly distributed in the retina. So the pig became the first model for retinitis pigmentosa, a cause of blindness. And four years ago, researchers created a pig model of cystic fibrosis² that, unlike mouse models, developed



T. J. Tabasco,
star of the show.

symptoms resembling those in humans.

Geneticist and veterinarian Eckhard Wolf at the Ludwig-Maximilian University in Munich, Germany, has exploited the similarity between the human and pig gastrointestinal system and metabolism — like us, pigs will eat almost anything and then suffer for it — to develop models of diabetes. One pig model carries a mutant transgene that limits the effectiveness of incretin, a hormone required for normal insulin secretion³. Mice with the transgene developed unexpectedly severe diabetes, but the pigs have a more subtle pre-diabetic condition that better models the human disease. “This shows the importance of using an animal with a relevant physiology,” says Wolf.

Pig models are now being developed for other common conditions, including Alzheimer's disease, cancer and muscular dystrophy. This work will be enriched by the discovery, reported in the genome paper, of 112 gene variants that might be involved in human diseases. Knowledge of the genome is also allowing scientists to try to engineer pigs that could be the source of organs, including heart and liver, for human patients. Pig organs are roughly the right size, and researchers hope to create transgenic pigs carrying genes that deceive the immune system of recipients into not rejecting the transplants.

Back on the farm, early knowledge about the pig genome led to the discovery in 1991 of a gene involved in porcine stress syndrome, in which the stress of overheating, being moved or even having sex causes the animals to die suddenly⁴. It then became possible to test for the ►

► gene and select pig stocks free of it.

Having the full genome should also help investigators to breed out susceptibility to porcine reproductive and respiratory syndrome (PRRS), a viral disease costing the US pig industry US\$600 million per year. The PRRS Host Genetics Consortium, a network of US research groups, has identified a region on one chromosome that affects levels of virus in the blood during infection⁵. Archibald, who works on PRRS, says that the high-quality genome

sequence should help investigators zero in on the genes responsible.

But the pig genome is not just about applications. Lead co-author Martien Groenen, a genome researcher from Wageningen University in the Netherlands, has resequenced the genomes of scores of different strains of wild and domestic pigs, and used the information to show that the pig was domesticated independently in Asia and Europe. He has also started to work out which genes were involved in the

selection of desired traits — such as a longer spine to give more bacon — on different continents. “It’s curiosity-driven research, but it may also help animal breeders in the future,” he says. ■

1. Groenen, M. A. M. *et al. Nature* **491**, 393–398 (2012).
2. Rogers, C. S. *et al. Science* **321**, 1837–1841 (2008).
3. Renner, S. *et al. Diabetes* **59**, 1228–1238 (2010).
4. Fujii, J. *et al. Science* **253**, 448–451 (1991).
5. Boddicker, N. *et al. J. Anim. Sci.* **90**, 1733–1746 (2012).

BUSINESS

Investment relief for biotech sector

Public markets provide cash injection for struggling field.

BY HEIDI LEDFORD

Robert Forrester gets a little giddy when he talks about the day his company went public. The otherwise understated chief operating officer of Verastem, a small biotechnology company developing drugs to target cancer stem cells, chuckles and bounces in his chair as he recounts key strategic decisions along the way to the company’s initial public offering (IPO) on 26 January, which raised US\$55 million.

Until recently, Verastem’s IPO would have stood little chance. Few biotech companies have braved an IPO in the years since the global recession hit, and those that did often took a beating in the public markets. Venture capitalists began to pull out of the sector. Colleagues scoffed when Forrester told them that Verastem, a young company in Cambridge, Massachusetts, with no clinical data was going public. “Many people said, ‘you’ve got to be kidding,’” he recalls.

But the IPO drought may be ending. This year has seen 12 biotech IPOs, and others are in the pipeline. So far, this has pumped some \$800 million into the sector, according to Renaissance Capital, an IPO-research company based in Greenwich, Connecticut. And biotech stocks are doing well — the NASDAQ Biotech Index has outperformed the NASDAQ Composite Index for the past 20 months (see ‘Bullish on biotech’). “If this trend holds, it could be great news for the sector,” says Josh Lerner, who studies venture capital at Harvard Business School in Boston, Massachusetts.

Restoring access to the public markets — particularly for young companies that have few fund-raising options left — can give companies

the capital they need to expand research programmes, hire more researchers or even just survive. It can also grant them access to ‘generalist’ investors who do not specialize in health care. “Public investors who may have been out of biotechnology for the past couple of years have started to move back in,” says James Healy, a general partner at venture-capital firm Sofinnova Ventures in Menlo Park, California.

Observers credit several factors for the rising investor confidence in biotech. Large pharmaceutical firms eager to restock drug pipelines are gobbling up smaller firms at high prices. Biotechnology companies have celebrated several high-profile successes in the past 18 months, with the US Food and Drug Administration approving groundbreaking drugs such as vemurafenib, a genetically tailored drug for advanced melanoma whose prowess in knocking out tumours is matched by its jaw-dropping price tag — more than

\$50,000 for six months of treatment.

Investors may also be drawn to the sector because of the poor performance of other industries, which are suffering more directly from the sluggish US economy, says Eric Schmidt, an analyst at investment bank Cowen and Company in New York. “Biotech earnings tend to grow independently of the economy, unlike electronics or consumer products,” says Schmidt. “Everybody needs medicine.”

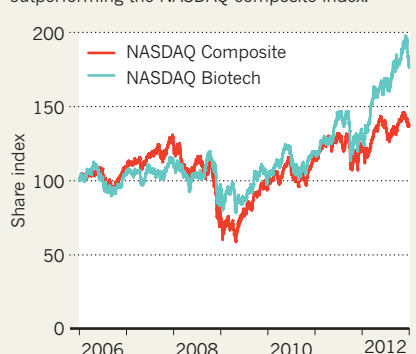
The wave of public investment could help to offset the dearth of venture capital. A survey released last year by the National Venture Capital Association, headquartered in Arlington, Virginia, showed that nearly 40% of venture capitalists had decreased their investment in biotech during the previous three years, put off by the long timelines and high risks of drug development. Several prominent health-care funds have closed altogether. “Biotech is a money-eating machine,” Lerner says. “The need for capital is so large, and given what’s happened to venture capital, having alternatives is important.”

Nowhere is that need greater than in young companies, the riskiest of all biotech investments, which have been among the hardest hit by the drop in venture funding. Verastem’s IPO money advanced the company’s business plan by two years; it should begin phase II trials of its leading compound by mid-2013, Forrester says.

But IPOs are not necessarily the answer for all struggling biotech ventures, cautions Brian Atwood, a managing director of Versant Ventures, a venture-capital firm in Menlo Park. He notes that many of the companies that pulled off IPO triumphs this year are unusual in some respect. Kythera Biopharmaceuticals of Calabasas, California, for example, is particularly appealing to investors because patients will have to pay out of their own pockets for its leading product — a fat-fighting injection designed to shrink double chins — rather than relying on health insurance and its accompanying cost controls. And Verastem’s Forrester can barely utter a sentence without referencing the company’s executives and scientific advisory board: a who’s who of Boston’s biomedical glitterati. Healy agrees: “It’s a higher-quality set of companies that have recently gone public compared with those that may have gone public five years ago.” ■

BULLISH ON BIOTECH

An index fund of biotech companies is outperforming the NASDAQ composite index.



SOURCE: NASDAQ

FUNDING

Berlin aims to create research powerhouse

German government finds a creative way to fund universities.

BY QUIRIN SCHIERMEIER

Berlin is poor but sexy," Klaus Wowereit once famously said of Germany's capital, where he has served as mayor since 2001. Now he is hoping that more money will buy it a little extra love, at least from biomedical researchers. Last week, he proudly announced an opulent deal to boost the city's scattered health-research base, sweetened with a hefty chunk of federal funding. Those in charge of the new Berlin Institute of Health (BIH) that was created by the deal believe that it could rival research powerhouses in the United States and Britain.

It also offers a model for circumventing the country's long-standing restrictions on federal funding of universities, a rule that baffles many outsiders. "That's the way forward, no matter what critics might say," says Wolfgang Herrmann, president of the Technical University of Munich, one of Germany's highest-ranked research universities.

The BIH will ally the Charité, Berlin's largest university clinic, with the Max Delbrück Center for Molecular Medicine (MDC), a national biomedical research centre. Over the next five years, the institute will receive more than €300 million (US\$380 million) in extra funding, 90% of which will come from federal budgets, with the remainder coming from the city and private sources. From 2018 onwards, the federal government will permanently support the BIH with an annual sum of €80 million. "That's a marvellous windfall," says Walter Rosenthal, the MDC's scientific director. "It could help us set up in Berlin something equivalent to the Howard Hughes Medical Institute in Maryland — it's a once-in-a-lifetime opportunity."

The MDC researches basic molecular medicine, but has little direct access to patients, so teaming up with the Charité should help to speed its discoveries to the clinic. Under the auspices of the BIH, molecular biologists and clinical researchers will join forces to tackle a spectrum of illnesses, from cancer to cardiovascular disorders, and neurodegenerative diseases such as Parkinson's and Alzheimer's. The BIH will complement Germany's six national health-research centres, which have narrower focuses. Work at the new institute, spread across Berlin's health campuses, could start next spring. Research coordinators at the Charité

and the MDC have put together a preliminary research plan, which an international evaluation team will review by March 2013.

A significant portion of the start-up cash, says Rosenthal, will be used to equip BIH groups with state-of-the-art sequencing, mass spectroscopy and bio-imaging technologies, and to expand the MDC's bioinformatics capacity. The newly equipped labs and secure long-term funding should lure biomedical researchers from around the world. Over the next 8 years, up to 60 collaborative groups could be set up, Rosenthal says.

"Recruitment is the bottleneck when it comes to turning Berlin into a health-science hub of truly international rank," says Claus Scheidereit, an oncologist who coordinates cancer research at the MDC. "If we can get over that, we can start to think really big."

The BIH is the German government's latest attempt to inject national funds into universities, circumventing a highly federalized system in which state governments jealously guard their responsibility for universities. The gov-

"It could help us set up in Berlin something equivalent to the Howard Hughes Medical Institute."

ernment's acclaimed 'excellence initiative', for example, invited German universities to compete for federal top-up grants and has so far generated thousands of new science jobs.

Collaborations that resemble the BIH, linking national research centres and universities, have also been forged in Karlsruhe, and between Jülich and Aachen.

A proposed amendment to Germany's constitutional law would allow the federal government to co-finance universities permanently, but it is unlikely to win the required two-thirds majority in parliament. Its opponents fear that the bill would allow Berlin to dictate national policy in areas, particularly secondary education, in which the states enjoy near-total control.

Scientists maintain that Germany's federalized university system is hampering the creation of national research hubs such as those in neighbouring Switzerland. "Why shouldn't the Technical University of Munich become Germany's federal institute of technology?" says Herrmann. "Because the rules don't allow it? If it helps our country, we should change them sooner rather than later." ■



FERTILE MIND

BY TRISHA GURA

Jonathan Tilly defied decades of dogma by suggesting that women can make new eggs throughout their lives. Now some of his critics are taking a second look.

Jonathan Tilly likes to gauge the significance of his work by the hair on the backs of his arms. “Look at it standing up,” he says, thrusting out his forearm on a mid-August afternoon. A reproductive biologist at Massachusetts General Hospital in Boston, Tilly was explaining a procedure to retrieve stem cells from the ovaries of a sterile woman. This experiment, he hopes, will help to quell criticism of his most controversial claim: that ovaries have the potential to make eggs indefinitely. This defies the long-held dogma that female mammals are born with all the oocytes (precursors to eggs) they will ever produce, a population that dwindles with age and is exhausted at menopause.

Tilly first challenged that doctrine in 2004, in a paper¹ suggesting that the oocytes in mouse ovaries are being replenished by stem cells. If properly understood, such cells could be harnessed to generate fresh eggs for women with fertility problems, or even achieve a goal Tilly has been pursuing for 25 years: delaying or halting menopause. “The hairs are still up,” Tilly says. It “happens every time I think about that experiment”.

He has since published a parade of headline-grabbing papers, culminating this year in a report² that he had isolated the elusive stem cells from human ovaries and coaxed them to develop into bona fide oocytes. But his work has been dogged by doubt. Some researchers question his methods and reasoning. Others have tried, and failed, to repeat his experiments. Tilly “always makes what I call ‘big satellites’, something tremendous in the sky,” says molecular biologist Kui Liu at the University of Gothenburg in Sweden. “He exaggerates,” Liu says, and produces a “big press release”. “A few years later, people realize, ‘Oh, not right.’”

Tilly says he has weathered a lot of attacks. “When I made the decision to pursue this, it was out of pure excitement that we found something that could revolutionize the field. It never even crossed my mind that it would be

so negative and so nasty. And it really is negative and nasty.”

But now the stand-off of mistrust, and sometimes open contempt,

has taken a strange twist. Two of Tilly’s most vociferous critics have become his collaborators: one serving on the board of advisers at his start-up company, OvaScience in Cambridge, Massachusetts; the other working directly with the stem cells that Tilly had isolated. “These cells are doing things *in vitro* that can really start to address scientific problems,” says Evelyn Telfer, a reproductive biologist at the University of Edinburgh, UK, who was doubtful of Tilly’s work in the past. “If we are really interested in the science ... then this is a great tool.”

A COUNTING PROBLEM

The ‘no new eggs’ doctrine has a long history. In 1951, the influential anatomist Solly Zuckerman, at the University of Birmingham, UK, performed an in-depth analysis of evidence available at the time. He concluded that none of it effectively countered a proposal from the 1870s stating that female mammals stop producing oocytes after birth³.

For the first 15 years of his career, Tilly focused mainly on programmed cell death, or apoptosis, and he was struck by the fact that no one had ever quantified the loss of eggs due to ovulation and natural oocyte death over time. So beginning around 1999, Tilly commandeered a microscope and mouse ovarian tissue in order to count the follicles, the cellular compartments in which oocytes develop, in mice at different ages. He found a mathematical imbalance: the number of degenerated follicles was three times higher than expected on the basis of the starting pool. If the mice were losing oocytes at this rate, their eggs should be depleted far sooner than they actually were. Something had to be replacing them, he concluded: stem cells were the likely culprit.

Few were willing to accept the idea. It took Tilly two years — and numerous rejections and revisions — to get the data published in *Nature*, in 2004. Controversy ensued over his methods as well as his conclusions. One critique said, for example, that it was “alarming” that Tilly used the rate of follicle disappearance in one mouse strain to calculate loss for another⁴.

Tilly dropped most of his apoptosis work and steered his entire lab towards proving the existence and functionality of these stem cells. “You are sort of standing on the precipice

wondering whether or not you should make the jump,” he says. “Getting the 2004 paper published was for me the jump, because there was no turning back at that point.”

A year later, Tilly reported that he had identified the source of these putative cells: bone marrow⁵. When he transplanted either marrow or blood from healthy mouse donors into sterile mice, the animals could produce cells that looked like oocytes. But he could not yet fertilize the resulting eggs and create embryos — the true test of an egg stem cell.

At least six groups challenged the bone-marrow finding. In one critique⁶, a group led by Telfer wrote that none of Tilly’s experiments had successfully been replicated, and that the results could be interpreted in other ways. Critics also asserted that Tilly was overreaching, particularly in media interviews.

In *The Boston Globe* in 2005, for example, Tilly is quoted as saying: “They’re your own cells; you don’t need anybody’s approval. They go right into your blood supply and go right to your ovaries, where they mature into eggs.” David Albertini, a reproductive biologist at the University of Kansas Medical Center in Kansas City, calls such claims outrageous: “A lot of us reproductive biologists feel that this is a frank travesty that has falsely raised the hopes of many women.”

Tilly defended his comments and challenged his peers to go back to their labs and reproduce his experiments. Several did. In 2006, stem-cell biologist Amy Wagers at Harvard University in Cambridge, Massachusetts, and her collaborators stitched together the circulatory systems of two mice⁷. One, the donor, expressed green fluorescent protein (GFP) in its cells. The other did not. The scientists found that although green, glowing, blood-borne cells could infiltrate the ovaries of the recipient mice, these cells acted like blood cells, not oocytes.

Tilly, in response, performed a similar experiment, showing that mice sterilized by chemotherapy could give birth after a bone-marrow transplant⁸. But the babies did not express GFP, indicating that the eggs from which they were derived came from the recipient, not the donor. Tilly argued that the bone marrow either protected existing oocytes or revived oocyte formation, but critics argued that the chemotherapy probably didn’t kill off all the recipient’s oocytes in the first place.

SHANGHAI SURPRISE

With little independent replication of his work, Tilly was standing alone through much of the fray. Then, in 2009, Ji Wu at Shanghai Jiao Tong University in China and her colleagues reported that they had isolated from mice what she called “female germline stem cells” — not from bone marrow, but from ovarian tissue⁹. When her team transplanted the cells into chemotherapy-treated female mice, they developed into mature oocytes, then fertilizable eggs and, the clincher, healthy pups.

➔ **NATURE.COM**
For a podcast
about this story,
visit:
go.nature.com/6rjbst

Although previously sceptical of the prospect, Telfer says that when she read Wu's paper in 2009 she paused, thinking "there must be something in this". She had met Tilly at the bar during a scientific meeting the year before and they talked about their differences. After Wu's paper, the two co-authored a commentary that articulated something like a truce¹⁰. "Although these findings do not establish that oogenesis occurs in adult females under physiological conditions," Tilly and Telfer wrote, "they strongly support the existence of [germline stem

cells] in adult mouse ovaries. If equivalent cells can be found in human ovaries, stem-cell-based rejuvenation of the oocyte reserve in ovaries on the verge of failure may one day be realized."

Many set out to replicate Wu's results, including Tilly. And in February, he reported the isolation of what he called "oogonial stem cells" from human ovaries². By injecting the cells into human ovarian tissue transplanted into mice, he was able to generate both follicles and what seemed to be mature oocytes (see *Nature* **483**, 16; 2012).

"So now we have two different labs, using conceptually a similar protocol, and both groups got confirmatory data," he says. "We felt at that point there should be no more debate."

But there was. Critics soon began pointing out a problem shared by both the teams' approaches. Each identify their respective stem cells using antibodies meant to bind a cell-surface protein, a common technique in cell biology. But the protein they target, called vasa, normally sits inside the cell, not on its surface. "There are a lot of people in the field struggling to understand how this can possibly work," says Patricia Hunt, a reproductive biologist at Washington State University in Pullman.

Tilly says that although mature oocytes do not express vasa on their surface, his cells — which are a cross between embryonic precursors and full-blown oocytes — do. Vasa, he says, becomes non-detectable on the cell surface as the cells mature into eggs. But, he adds, "We don't have any proof of that yet."

Liu in Sweden says that he initially believed Wu's paper when it came out. But his group could not repeat the technique. To bypass the cell-surface problem with vasa, Liu used an approach that tracks the protein inside the cells¹¹. He was able to extract ovarian, vasa-expressing cells, but none of them underwent division — a major criterion for stem cells.

Wu contends that her cell-isolation technique is not easy to perform and invites scientists to come to her lab to learn it. She adds that Liu's group "didn't use our protocol of isolating the cells. So how to compare?"

In fact, Tilly says that his lab had trouble repeating Wu's protocol, too. Eventually, his team retrieved cells but "found consistent oocyte contamination". He had to modify the protocol to retrieve the mouse and human oogonial stem cells, and they differed in size from those Wu had isolated. Wu says that her cells and Tilly's are probably "subtypes" of each other and that there is still "a lot of work to do" to figure out exactly how they are related.

Telfer, meanwhile, has begun to collaborate with Tilly. After going to Boston in 2011 to observe his human stem cells, she was impressed, and took a sample back to Scotland. Her team had worked out a culture system using mouse and cow tissues to grow egg precursor cells into fertilizable eggs entirely outside the body. With Tilly's cells she needed to adapt the technique for use in humans. "The first experiments blew me away, just blew me away," she says. Tilly's cells, she found, grew rapidly into oocyte-like structures. "I spent the whole night trying to find another explanation other than new follicles had formed," she says. "And I could not come up with one."

Telfer has applied for permission from the UK Human Fertilisation and Embryology Authority to attempt to fertilize the cells; such experiments are forbidden using US federal funding. If successful, the technique to make fertilizable human eggs outside the body could eventually be disseminated to fertility clinics throughout the world.

The cells are also being used at OvaScience, which was founded in April 2011 and has secured US\$48 million in venture capital. The company is exploiting Tilly's cells in several ways. One aim is to rejuvenate egg cells from older women by adding fresh cytoplasm and mitochondria. The research builds on a controversial experimental fertility technique in which egg cells are injected with cytoplasm from another woman's eggs. The OvaScience approach would use mitochondria extracted from the mother's own oogonial stem cells, which Tilly says would be healthier than those from an ageing mother's eggs, and should skirt some of the ethical and safety questions raised by using donor mitochondria. OvaScience plans to begin clinical trials this year in collaboration with two Boston-based fertility clinics.

BOUNDARIES AND BACKLASH

Tilly's oogonial stem cells will also serve as a screening tool for new drugs that might block

or boost egg production. Such drugs might help reverse infertility or even help delay or halt menopause. Albertini still worries that such claims inflate hopes but, like Telfer, he is trying to keep an open mind. The prospect of new models for screening fertility drugs convinced him to join OvaScience's scientific advisory board. "There are a lot of things that I know and I do that could be helpful to them," he says.

The company is swiftly moving forward, steered by the team that guided Sirtris, based in Cambridge, Massachusetts, a biotech firm focused on anti-ageing therapies. In fact, it was a collaboration between Tilly and Sirtris's founder, David Sinclair, at Harvard Medical School in Boston, that sparked the launch of OvaScience.

He and Tilly are "mutual admirers", Sinclair says, explaining that they joined forces in 2009 to explore the idea that egg quality declines with age because older eggs lack enough energy to support fertilization. Sinclair offers his anti-ageing expertise and his experience of controversy; some of the initial results on which Sirtris was founded could not be replicated and have been a source of contention in the field (see *Nature* **464**, 480–481; 2010).

"It is an interesting team that Jon and I make," Sinclair says, "because the two of us push the boundaries of science. And both of us have encountered backlashes in doing so."

Still, despite his characteristic gumption and ebullience, Tilly seems to be burdened by the continual sparring. Although he's shifted much of his time to studying oogonial stem cells in the ovary, he still maintains that bone-marrow stem cells might also create new eggs. His critics disagree, and even if they accept the existence of oogonial stem cells, they still question whether such cells normally function to produce new eggs.

"The data provided so far don't support this concept," says Albertini. Tilly maintains that these stem cells must be doing something in the body. But in exasperation, he is willing to concede that it may not matter in the clinic. "If you could take these cells outside the body, and get them to make a functional egg that can make a normal healthy baby, what do you care about the physiology?" ■

Trisha Gura is a freelance writer in Boston, Massachusetts.

1. Johnson, J. et al. *Nature* **428**, 145–150 (2004).
2. White, Y. A. R. et al. *Nature Med.* **18**, 413–421 (2012).
3. Zuckerman, S. *Recent Prog. Horm. Res.* **6**, 63–109 (1951).
4. Gosden, R. G. *Hum. Reprod. Update* **10**, 193–195 (2004).
5. Johnson, J. et al. *Cell* **122**, 303–315 (2005).
6. Telfer, E. E. et al. *Cell* **122**, 821–822 (2005).
7. Eggan, K. et al. *Nature* **441**, 1109–1114 (2006).
8. Lee, H. J. et al. *J. Clin. Oncol.* **25**, 3198–3204 (2007).
9. Zou, K. et al. *Nature Cell Biol.* **11**, 631–636 (2009).
10. Tilly, J. L. & Telfer, E. E. *Mol. Hum. Reprod.* **15**, 393–398 (2009).
11. Zhang, H. et al. *Proc. Natl Acad. Sci. USA* **109**, 12580–12585 (2012).



QUANTUM LEAPS

Fully fledged quantum computers are still a long way off. But devices that can simulate quantum systems are proving uniquely useful.

BY GEOFF BRUMFIEL

When high-energy physicists announced in July that they had found the long-sought Higgs boson — their biggest find in decades — the thousands of individuals involved rightly held their heads high. But in some sense, they had already been beaten to the prize.

Months earlier, a team of nine physicists had taken a rarefied vapour of rubidium-87 atoms, cooled it down to very near absolute zero and used lasers to arrange the atoms into a tiny grid. The physicists then tweaked the temperature until the atoms neared a critical ‘phase transition’ — a point between two different behaviours, such as liquid water and solid ice. Monitoring their grid in this in-between region, the researchers saw an unusual wave of energy that appeared momentarily and then died away¹. Mathematically speaking, this behaviour was the same as the appearance and decay of a Higgs particle inside a particle collider.

“Obviously, it’s not at all the Higgs particle,” says Immanuel Bloch, the researcher who led the study at the Max Planck Institute for Quantum Optics in Garching, Germany. If nothing else, this particle moved in only two dimensions, whereas the Higgs moves in three. But the experiment is still helpful for particle physicists, says Bloch, because it gives them a new way to explore and test the complex quantum field theories that underlie the Higgs.

This experiment also put Bloch and his team at the vanguard of the rapidly growing field known as quantum simulation. The idea, broadly speaking, is to use orderly systems such as a grid of atoms to model much more complicated things — new particles, for example, or high-temperature superconductors. The behaviour of such systems cannot be derived by hand, and even the world’s fastest supercomputers can’t model them.

Quantum simulators are the lesser sibling of an idea in physics known as quantum computers, which have been touted for more than three decades as a way to do everything from complex modelling to code-breaking. What the simulators and computers share is an ability

to operate by the rules of quantum mechanics. Where they differ is in computational power: quantum computers are general-purpose machines able to carry out any possible algorithm, whereas quantum simulators have to be tailored specifically for the problem at hand. Current-generation simulators are also tough to control, and they may not be able to tackle every problem. Nevertheless, the simulators are much easier to build than quantum computers. And researchers say that the devices will soon be able to solve at least some quantum problems that can’t be tackled in any other way.

NUTS AND BOLTS

The world of quantum physics is full of theorems, but one goes unwritten: if you want to get noticed, show that your idea came from Richard Feynman.

Feynman, the mid-twentieth-century’s greatest theoretical physicist, came up with the idea of quantum simulation in 1981 when he was asked to deliver a keynote speech at the Massachusetts Institute of Technology (MIT) in Cambridge². He decided to talk about how physics might be simulated with computers and got straight to the core of the problem: computers run on certainty, but at a fundamental level, nature deals in probability. According to the laws of quantum mechanics, he knew, particles very rarely exist in one state or another, but instead live in a ‘superposition’ of two states at once. When observed, the paradox resolves itself according to the laws of statistics. For example, an electron’s ‘spin’ may orient itself in one direction half the time, and in the other direction for the other half.

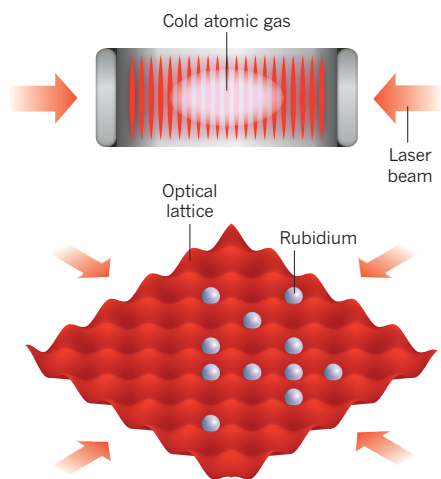
It is not hard to program a normal computer to model the probabilistic behaviour of that one electron, said Feynman. But particles do not live in isolation, and in quantum systems their probabilities are linked, or ‘correlated’. These correlations mean that every combination of particle states must be computed separately, and this creates an exponential rise in complexity. A system with three electrons has eight possible configurations, with eight probabilities to

QUANTUM BOARD GAMES

The set-ups of quantum simulators are different, but the concept is the same: first take atoms, ions or electrons, cool them to cryogenic temperatures and arrange them in an orderly grid. Then tune the interactions on the grid to mimic a more complex material.

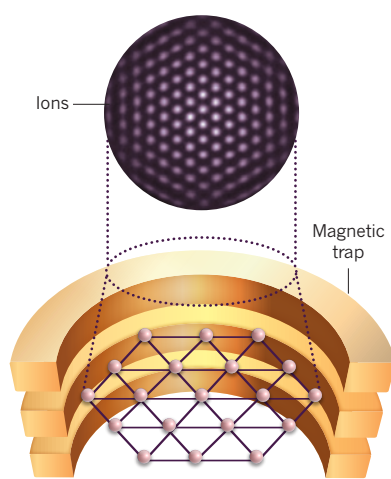
COLD ATOMS

Rubidium atoms are held in place by criss-crossed laser beams, which can also be used to tweak individual particles. A single pair of lasers holds the atoms in a one-dimensional column (top), whereas two pairs hold them in a grid (bottom). Some excitations in the grid system behave like the Higgs particle.



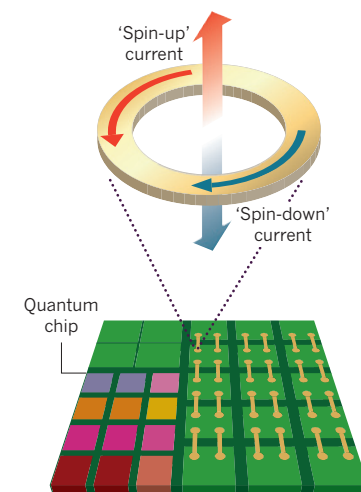
TRAPPED IONS

A combination of electric and magnetic fields trap charged, ionized atoms in an orderly grid. The ions wiggle and rotate in a way that mimics the interactions of quantum magnetism — a phenomenon that can't be simulated in classical systems.



SUPERCONDUCTING LOOPS

A quantized loop of current can flow clockwise, anticlockwise or in a superposition of both in a superconducting circuit (top). An array of such loops (bottom) can be manipulated to simulate various quantum systems — and perhaps even biological processes such as photosynthesis.



compute; 300 electrons create as many configurations as there are atoms in the known Universe.

Feynman spent most of his lecture trying to find a way out of this conundrum. It is not easy using ordinary computers, he concluded, but there is another possibility: build a computer that thinks in terms of probabilities. This quantum imitator, as he called it, would look a lot like whatever system you were trying to model. It wouldn't need to crunch every outcome, but instead would simply recreate the range of probabilities. Rather than delivering one solution, the imitator would deliver many, and the likelihood of each answer would create a probabilistic picture of how the complex system behaves. Feynman didn't do the maths, but he did conclude that almost any quantum system "can be simulated in every way, apparently, with little latticeworks of spins and other things".

At the time of Feynman's talk, the little lattices of which he spoke didn't exist. Quantum systems are extremely fragile, in the sense that almost any interaction with the outside world will destroy the delicate correlations. It has taken 30 years to develop the technology required to keep the particles isolated enough to finish the simulation unimpeded, yet interactive enough to let physicists extract the answer. But there are now several options. Bloch's group uses neutral atoms, other teams are combining electric and magnetic fields with lasers to trap ions of lighter atoms, such as beryllium. A third technique involves controlling eddies of current inside superconducting microcircuits, and a fourth uses quantum particles of light — photons — moving through microscopic waveguides (see 'Quantum board games').

All these techniques are rapidly increasing in their capabilities. In April, a group led by John Bollinger at the National Institute

of Standards and Technology in Boulder, Colorado, unveiled a two-dimensional system of hundreds of trapped ions that could simulate a form of quantum magnetism³. The simulator seems to work well for weak fields of the sort that can already be modelled on classical computers, says Bollinger. Now, with some modifications, he hopes to simulate strong magnetic fields, which are beyond the reach of even the most powerful supercomputers.

Bloch, meanwhile, is considering applications beyond the Higgs for a neutral-atom simulator. For example, the rubidium atoms in his lattice might be used to model a complex class of materials called high-temperature superconductors. These 'high- T_c ' materials can

conduct electrons with no resistance at temperatures much higher than conventional superconductors can — but for decades nobody has been able to understand why. Theorists have developed a number of competing models to explain the behaviour, but haven't been able to test them: the electrons in the superconductors are just too difficult to isolate and study. So Bloch

wants to use atoms as surrogates. By changing the intensity of the criss-crossing laser beams, atoms can be made to tunnel from one point in the lattice to another in a way that mimics the motion of electrons through the atomic lattice of a high- T_c material. At least some theories of high- T_c superconductivity should be checkable with Bloch's set-up.

Quantum simulators might even be able to model non-quantum problems, such as protein folding, that still require huge amounts of computing power to decipher. A group at the Canadian company D-Wave Systems in Burnaby and at Harvard University in Cambridge, Massachusetts, recently did just that by mathematically mapping

"IN A QUANTUM COMPUTER YOU HAVE TO MAKE SURE THAT NO PARTICLE MAKES A MISTAKE."

the folding problem onto a quantum system of 128 loops of current spinning on a superconducting chip⁴. Each loop could spin clockwise, anticlockwise or in a superposition of both directions simultaneously. The performance of the system wasn't great; in one of its protein-folding problems, it found the correct, experimentally determined, answer just 13 out of 10,000 times. Still, says Alán Aspuru-Guzik, a theoretical chemist from Harvard and co-author of the paper, "it's remarkable to me that it was possible to do it" at all.

GOAL CHANGE

Despite all the technical progress, however, the existing simulators are at best a limited approximation of Feynman's original vision — a fully fledged quantum computer that is 'universal', or able to execute any quantum algorithm and simulate any conceivable quantum system. Researchers have been exploring the potential applications of such a device ever since Feynman described it. Arguably the most important one came in 1994, when mathematician Peter Shor, now at MIT, laid out an algorithm that would allow a quantum computer to function as a powerful code-breaking machine⁵. Other quantum algorithms have followed, drawing many scientists (and several intelligence services) into the quest for quantum computing and sparking widespread efforts to create such a machine.

Yet building a powerful, universal quantum computer has proven to be a tough task. A true Feynman computer would be able to control thousands or millions of atoms at once, but most of the current systems face a trade-off between size and control. Bloch, for example, can hold as many as hundreds of thousands of atoms in his laser lattice, but he can't then set their quantum states individually. Other researchers have more control over individual atoms, but their systems, which use trapped ions of beryllium, can manage only a handful of atoms with exquisite precision. On top of this comes the omnipresent problem of disruptions from the outside world, which ruin delicate quantum states: even the tiniest bump will create a computational error.

With current systems so far from the ideal, quantum simulators have come to be seen as less of a stepping stone, and more of a goal in their own right. Simulators do not need to be as large as computers, and, crucially, because the answer is encoded as an average across all their atoms, they are believed to be tolerant of the outside disruptions. "In a quantum computer you have to make sure that no particle makes a mistake," says Ignacio Cirac, a theorist at the Max Planck Institute for Quantum Optics. "In a quantum simulation, if you have 100 particles and one of them is wrong, then 99 are still right."

Some see parallels to the middle of the last century, when scientists such as Vannevar Bush were experimenting with 'analog' computers made from resistors and capacitors. The machines were tailored to specific problems or to a class of problems, and could perform a simple set of operations on an input signal. Some of the devices could even perform mathematical calculations. In retrospect, they seem puny compared with digital computers, which use programmable combinations of transistors to perform practically any program. But they were fast, robust and valuable for applications that matched their architecture, says Seth Lloyd, a theoretical physicist and engineer at MIT. They were particularly good at controlling machinery, for example. "All the control circuits in the Saturn moon rocket were analog," Lloyd says.

Like analog computers, quantum simulators are closely tied to their constituent parts, and are less flexible than a true quantum computer. But Lloyd thinks that they might yet find their 'Moon shot' in problems of quantum complexity. For example, as microprocessors

shrink and new materials are engineered at a molecular level, quantum effects become more and more important. That, in turn, will lead to a dramatically growing need for quantum modelling that allows designers to understand and predict the materials' behaviour. At least some of those needs are going to be met by quantum simulators, Lloyd predicts. "What seems to be happening is that quantum simulators work on a variety of special cases," he says, "and the number of cases seems to be growing rather rapidly."

Aspuru-Guzik has one such process in mind: photosynthesis. When light strikes a leaf, it creates a pair of negative and positive charges that travel long distances to reaction centres, where they are used to make energy for the plant. The charge pairs may travel according to the rules of quantum mechanics: some researchers think that the collective wavefunction of the pairs spreads out across the light-absorbing chromophore molecules inside the leaf, allowing the pairs to move more efficiently than they would classically (see *Nature* **474**, 272–274; 2011).

Aspuru-Guzik and others think that a simulator could help them to pin down exactly how this happens. Photosynthesis is what Aspuru-Guzik calls a "dirty quantum system" — that is, it contains both quantum and classical elements. A little matrix of superconducting current loops might be perfect for modelling it, he argues, because the loops, too, are subject to noise from the outside world. It still wouldn't

be easy, however: Aspuru-Guzik estimates that something such as photosynthesis would require hundreds of quantum bits to simulate, and those systems, he predicts, are at least a decade away.

The ambitions of the scientists developing quantum simulators are considerably more modest. Most are starting their systems out on models that can be calculated with conventional supercomputers to prove that their simulators produce reliable results. Gradually, they plan to push their atoms, current loops or other little units to the point at which the supercomputers can no longer cope.

At that point, "the model that we're able to implement might not even correspond to a real material, but in a sense, who cares?", says Chris Monroe, a physicist at the University of Maryland in College Park. Even if they don't behave like a superconductor or a Higgs particle, the new systems may still be able to tell researchers a thing or two that their older machines can't. Eventually, Monroe and others believe that simulators will be tailored to model different things. Cold atoms, for example, might work best on superconductors, whereas ions could handle magnetism. Of course, there will still be quantum systems that are too tough for any set-up to tackle.

It may be a vision considerably less flashy than Feynman's universal quantum machine, yet within the physics community, quantum simulators are getting more attention than ever before. "Many physicists who sort of pooh-poohed the idea of quantum computing, especially ten years ago or so, they're now sort of embracing this," says Monroe. The systems may be less ambitious, but that may make them more achievable.

Lloyd puts it another way. "If life doles you quantum lemons, let's make quantum lemonade," he says. Simulators may not be as sweet as quantum computers, but "as long as the lemonade is tasty and refreshing, I think that's fine." ■

Geoff Brumfiel is a senior reporter at Nature.

1. Endres, M. *et al.* *Nature* **487**, 454–458 (2012).
2. Feynman, R. P. *Int. J. Theor. Phys.* **21**, 467–488 (1982).
3. Britton, J. W. *et al.* *Nature* **484**, 489–492 (2012).
4. Perdomo-Ortiz, A., Dickson, N., Drew-Brook, M., Rose, R. & Aspuru-Guzik, A. *Sci. Rep.* **2**, 571 (2012).
5. Shor, P. W. *SIAM J. Comput.* **26**, 1484–1509 (1997).

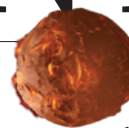
"THE MODEL THAT WE'RE ABLE TO IMPLEMENT MIGHT NOT EVEN CORRESPOND TO A REAL MATERIAL."

COMMENT

BIOTECHNOLOGY A call for more rigorous research into health impact of GM foods **p.327**

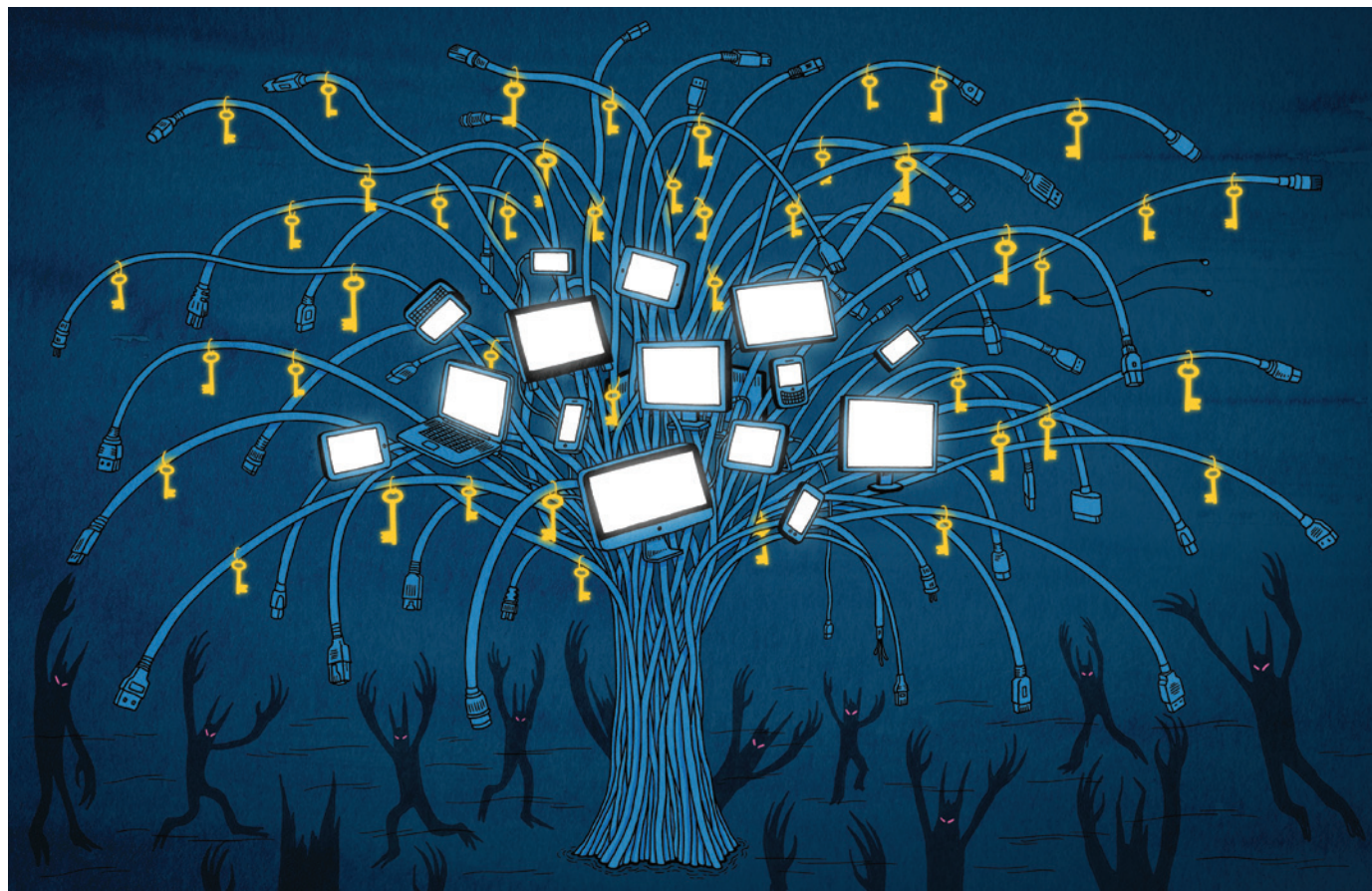
CULTURE Conservation in Italy 40 years on from UNESCO heritage list **p.328**

ASTRONOMY Extrasolar planets, sci-fi and Kim Stanley Robinson **p.330**



OBITUARY Edward Donnall Thomas, bone-marrow pioneer, remembered **p.334**

ILLUSTRATION BY ANDREW RAE



Secure the Internet

Software engineers must close the loophole used to intercept online communications, say **Ben Laurie** and **Cory Doctorow**.

In 2011, a fake Adobe Flash updater was discovered on the Internet. To any user it looked authentic. The software's cryptographic certificates, which securely verify the authenticity and integrity of Internet connections, bore an authorized signature. Internet users who thought they were applying a legitimate patch unwittingly turned their computers into spies. An unknown master had access to all of their data.

The keys used to sign the certificates had been stolen from a 'certificate authority' (CA), a trusted body (in this case, the

Malaysian Agricultural Research and Development Institute) whose encrypted signature on a website or piece of software tells a browser program that the destination is bona fide. Until the breach was found and the certificate revoked, the keys could be used to impersonate virtually any site on the Internet.

Fake certificates are used by hackers and governments to harvest online communications. In 2011, for example, a hacker based in Iran stole the signing keys from DigiNotar, a Dutch CA that declared

bankruptcy soon afterwards. The keys were used to impersonate sites such as Facebook and Gmail in Iranian dissidents' browsers, allowing all of their messages to be read.

Certificates allow the web to work. They secure transactions and allow users to enter credit-card numbers, share data across networks or chat in private forums. Without certificates, hackers could easily stop, corrupt or eavesdrop on these exchanges. But certificates are in trouble. As more authorizing bodies are added to browsers' lists of trusted CAs, and as governments, ►

► hackers and unscrupulous insiders weaken the Internet's security system, it is becoming virtually impossible to know whether a connection is legitimate.

Many Internet technicians, ourselves included, agree that it is time to fix the problem. What remains is the substantial hurdle of reaching consensus about how.

The international Internet Engineering Task Force is adding registration information to the (already overburdened) domain name system (DNS). But websites can still be taken over. The Electronic Frontier Foundation (where C.D. is a fellow), based in San Francisco, California, has proposed a cryptographic protocol called Sovereign Keys (SK) that would make it impossible for a third party to impersonate any website.

"It is becoming virtually impossible to know whether a connection is legitimate."

A third effort is under way, led by a team (including B.L.) at Google, based in Mountain View, California. This protocol — called Certificate Transparency (CT) — is similar to SK, but it includes an independent cross-checking system. Release dates have not been set for either protocol, but CT has the potential to be rolled out sooner, through regular software updates for Google's web browser, Chrome. We see it as a stepping stone to a more ambitious system, such as SK.

We call on browser vendors to support a shift to a more secure system. There are economic barriers — no one is likely to make money from shoring up the Internet. But the risks of ignoring this security loophole are too great.

CROSS-TALK

Before your browser connects to a website, it asks your local network's DNS server for the numeric address corresponding to the website's domain name. (For example, one of the addresses for www.facebook.com is 66.220.149.11.) But DNS is not secure — its communications with browsers are unscrambled, and they are easy to intercept. Anyone sharing your network can steal your credit-card information or passwords.

To scramble messages and keep them private, you need encryption. If a browser has received a cryptographic certificate signed by a CA, its address bar shows a key or padlock icon. All browsers have pre-installed lists of trusted CAs against which to check certificates.

CAs do a lot of due diligence before issuing a certificate, but they are fallible. Cryptographic methods can be used to spot forgery and tampering, but not to distinguish real certificates issued by diligent CAs from those issued by mistake or by a

CA that has been conned or taken over.

The proliferation of CAs is putting the entire Internet at risk. Governments are not responding to the problem — indeed, some policy-makers have shown a remarkable willingness to undermine online security for law-enforcement reasons. India's government, for example, is seeking weaker security for Skype and BlackBerry mobile devices¹. In 2011, US lawmakers proposed the Stop Online Piracy Act, which would require DNS providers to return false results when users try to connect to sites accused of facilitating copyright infringement².

Luckily, software engineers are in a position to fix the certificate loophole.

CERTIFICATE TRANSPARENCY

CT and SK rely on a type of record that uses cryptographic methods to prove that none of its past entries has been erased — an 'untrusted, verifiable, append-only log'. This log is based on a mathematical principle called a Merkle tree — a hierarchy of linked items, or 'leaves'.

In CT, each leaf is a certificate. For a particular node in the tree, a value can be generated in a few steps on the basis of the values of other nodes. For a tree with, say, a million leaves, a verifier would have to track only 20 nodes to confirm any particular leaf. Even a tiny alteration throws off the calculations entirely. With Merkle trees it is possible to prove efficiently that a particular leaf is in the tree without revealing the contents of the other leaves. It is also impossible to fake such a proof for a leaf that is not in the tree.

Merkle-tree logs are stored on a small number of computers, or log servers. Every time a CA generates a new certificate, it sends a copy to all the log servers, which then return a cryptographically signed proof that the certificate has been added to the log. Browsers could be pre-configured with a list of verified log servers (in addition to the list of CAs now installed).

Periodically — perhaps hourly — a number of 'monitor' servers contact the log servers and ask for a list of all the new certificates for which they have issued proofs. These monitors — operated by companies, banks, individuals and service providers — would discover any unauthorized certificates, just as credit reports alert people to cards or loans issued falsely in their names.

This process works only if the log servers are honest; here, auditor servers come in. Every so often, a browser sends all the proofs it has received to a number of auditors — anyone may act as one, because the logs are public. If a proof has been signed by a log server but does not appear in its log, the auditor knows that something is wrong. Within an hour of committing their first transgressions, rogue CAs and

log servers could be detected and removed from browsers' lists.

Even with modest uptake, CT will begin cleaning up the Internet immediately on roll-out. Blocking will improve as more organizations, browser vendors and users participate. Initially, browsers that adopt CT will not be able to block connections for which no proofs are offered. After a year, Chrome will be updated to warn users before establishing a secure connection without a proof. Later, it will not connect to any site without one. We hope that other browsers will follow a similar path.

Also based on Merkle-tree logs, SK is a more theoretical approach. Instead of requiring CAs and domain registrars, SK issues a private key for each website to only one holder. No one else may use that key, so misuse by a third party is not a problem. Anything unverified is blocked, so SK would foil attempts by governments to use domain seizures to censor content that they find to be objectionable.

LONG-TERM GAIN

Through systems such as CT and SK, software engineers can and should solve the certificate problem. After all, the Internet is international and independent; governments cannot mandate a solution.

There are some economic barriers to improving security, but it is a worthwhile investment. CAs have no short-term incentive to support these measures, but the CA network offers the best avenue for rolling out steps such as CT. Similarly, someone must run the log servers, even though doing so will not lead to direct economic gain. But the long-term stability and security of the Internet is good for business. In line with this view, Google will run some log servers — but others should as well, so that we can avoid putting all of our eggs in one basket. Browser vendors should also commit to supporting CT.

History tells us that those who seek new avenues of attack will eventually find them. But this troubling breach must be closed down now. ■

Ben Laurie is a visiting industrial fellow at the University of Cambridge, Cambridge CB2 1TN, UK. **Cory Doctorow** is a visiting senior lecturer in the Computing Department at the Open University, Milton Keynes MK7 6BJ, UK.
e-mail: ctnaturepaper@links.org

1. Anonymous. India to seek Interpol help to intercept encrypted data from BlackBerry, Gmail, Skype. *The Economic Times* (20 December 2011).
2. US House of Representatives, 112th Congress, 1st sess. bill no. HR 3261 (2011).

Competing financial interests declared; see go.nature.com/cdgvuun.

Bring more rigour to GM research

The latest furore over GM food highlights the need for good-quality research on highly sensitive topics, says **François Houllier**.

This autumn, a team of French researchers published results showing that rats fed with genetically modified (GM) maize (corn) died younger and showed more organ damage and tumours than usual. The team also observed similar effects in rats exposed to a combination of the GM maize and the herbicide it is designed to tolerate, and to the herbicide alone¹. Not surprisingly, these results sparked debate among the public and many in the scientific community.

The public concern is easy to comprehend. The images of tumour-ridden rodents included in the study tapped into existing controversies over the safety of GM crops going back several decades, especially in Europe. Scientific reactions were intense for the opposite reason: other research on exposure to GM foods has not shown such pathological patterns².

Since the paper was published, members of the scientific community have found weaknesses in the analysis, which, in their view, call the conclusions into question. But some damage may have been done: an opinion poll a few days after the paper's publication — although not specifically mentioning the study itself — showed that 79% of French people were worried about the possible presence of GM organisms in their food, compared with 65% in 2011 (ref. 3).

In my opinion, this episode highlights a major issue: there is a need for extra rigour in research whenever it tackles sensitive topics such as GM crops and food. Until science moves to the forefront, I believe the debate risks remaining mired in confusion and misinformation, no matter what improvements are made in public engagement.

Reactions to this latest GM study were reinforced by an unusual communication campaign. The authors informed a few journalists about the paper in advance, and asked them to sign an agreement saying that they would not interview outside experts until after the story had appeared in a French weekly magazine, *Le Nouvel Observateur*. In the following weeks, two books and a documentary based on the conclusions of the study were released. The effect was immediate: in my view, there was an initial wave of one-sided alarming news reports and increased distrust towards "the system".

The French government and the European Commission immediately asked national and European food-safety agencies to review the publication. Most have now released preliminary or final reports pointing out weaknesses.



Field trials of a GM grapevine rootstock were destroyed by activists in France in 2010.

These include a lack of relevant statistics — for example, on mortality and tumour incidence — resulting from the use of too few animals per group for long-term studies and a lack of plausible biological mechanisms for understanding the alleged effects. In my view, the paper seems to have failed to convince many in the scientific community, despite asking legitimate questions about long-term toxicity tests and the effects of a herbicide. Study author Gilles-Eric Seralini, of the University of Caen, has agreed that more animals would render the study more robust, but says that his findings are supported by many observations; he has also accused many detractors of conflicts of interest.

How do we address the questions about the impact of GM crops, and how do we prevent this kind of negative reaction? First, I believe that we need to publicly fund more risk-benefit analyses of GM crops. We also need more interdisciplinary studies of GM foods, especially on health impacts in animals and humans. A review² identified 24 papers featuring trials of feedstuffs containing various GM crops, in which the trials lasted more than 90 days or were done in more than one generation. By contrast, more studies have charted the environmental impacts of GM crops, including long-term, large-scale studies and meta-analyses (see, for example, refs 4,5).

Research into GM crops can be difficult. For example, at the French National Institute for Agricultural Research (INRA) in 2005, we launched a programme to test the environmental impact of a GM grapevine rootstock that was supposed to be resistant to grapevine fanleaf virus, which causes large economic losses. The project was funded only by public money; it did not aim to develop a commercial variety. There was a public consultation about it, moderated by specialist

social scientists^{6,7}, and stakeholders were transparently informed. Nonetheless, activists destroyed the crop in August 2010.

Second, research must always follow proper academic standards. In my opinion, any breach in the rigour and traceability of the scientific workflow — stating the research question and hypothesis; designing adequate experiments; using relevant data analysis and modelling techniques; allowing outside experts to comment on the results — could, I fear, lead to a lack of trust.

Publishing a paper is not the end of the story. New data and results should be tested by the scientific community until a convergent corpus of evidence is reached by independent teams. I believe that publicizing and sharing raw data and disseminating new methods are thus extra crucial stages. The more unexpected the results, the more rigorous this workflow should be.

Third, the distinctions between scientific research, risk assessment and risk management must be clearly articulated. This is vital for public trust in the long term. The GM maize case has generated the feeling that research organizations should do risk-assessment trials. I disagree: they should focus on elaborating and testing new methods — such as how we can use metabolomics to get early predictors of metabolic impacts — and on dissecting underlying biological mechanisms. In my view, risk assessment should remain within the province of dedicated agencies using specific guidelines and impartial procedures, albeit informed by the best research, as happened with the chemical bisphenol A^{8,9}.

As scientists, we must champion the multiple concerns of society, even when they make a contradictory call for more innovation as well as more precaution. ■

François Houllier is president and chief executive of INRA, Paris, France.
e-mail: francois.houllier@paris.inra.fr

1. Seralini, G.-E. *et al. Food Chem. Toxicol.* **50**, 4221–4231 (2012).
2. Snell, C. *et al. Food Chem. Toxicol.* **50**, 1134–1148 (2012).
3. IFOP. *Les Français et les OGM*. Available at <http://go.nature.com/upf1cx> (in French).
4. Lu, Y. *et al. Nature* **487**, 362–365 (2012).
5. Marvier, M. *et al. Science* **316**, 1475–1477 (2007).
6. The Local Monitoring Committee, Lemaire, O., Moneynon, A. & Masson, J. E. *PLoS Biol.* **8**, e1000551 (2010).
7. Joly, P. B. & Rip, A. *Nature* **450**, 174 (2007).
8. Arnich, N. *et al. Int. J. Hyg. Environ. Health* **214**, 271–275 (2011).
9. Vandenberg, L. N. *et al. Endocr. Rev.* **33**, 378–455 (2012).

V. KESSLER/REUTERS/CORBIS



The Colosseum in Rome: a much-needed €25-million conservation project is set to begin next month.

CONSERVATION

Shoring up the wonders

Forty years on from UNESCO's world heritage convention, **Alison Abbott** contemplates the state of Italy's vast legacy.

Even in the brightest sunshine, Mount Vesuvius casts a threatening shadow over Naples in southern Italy. Residents live in fear of the volcano, whose murderous eruption in AD 79 propelled lava and ash over surrounding towns, including Pompeii and Herculaneum, burying them.

For centuries, the ancient towns remained safely sealed from the elements; they were rediscovered only in the eighteenth century. Archaeological excavations since then have revealed much about life in Roman times, but Pompeii in particular dominates the public's imagination. The 66-hectare site, two-thirds of which has been excavated, receives more than 2 million visitors a year. Many Neapolitans make their living thanks to the tourist industry created by the catastrophe.

But a new shadow has fallen on the sites. The collapse of some structures during the past few years — including Pompeii's Schola Armaturarum or 'House of the Gladiators' in November 2010 — has raised questions

about whether Italy is taking good enough care of its considerable cultural heritage. Concerns have been inflamed by a well-publicized series of calamities, small and large, at several other sites in Italy, including stones falling from the walls of the Colosseum a year ago.

Italy has the largest number of entries of any country on the World Heritage List, which was created on 16 November 1972 under the UNESCO Convention Concerning the Protection of the World Cultural and Natural Heritage. Along with Italy's place on the list comes moral pressure to safeguard its heritage — artefacts, artworks and architectures from the Etruscan and Roman periods, through the Renaissance and up to the twentieth-century dictatorship of Mussolini, which put an end to Italian glory. Minor amphitheatres in remote towns like Cassino and specialized scientific collections such as the University of Pavia's eccentric hoard of pathological specimens are considered no

less important than better-known items.

It is often forgotten just how much Italy is doing right where its heritage is concerned. Many important sites and artworks are in fine shape — for example, the painstakingly restored *Last Supper* by Leonardo da Vinci in Milan. Between 1977 and 1999, under the guidance of Pinin Brambilla Barcilon, conservators used techniques such as chemical analysis of different layers of the fresco in microscopic core samples and infrared reflectoscopy to see below the surface of the fresco without harming it. Indeed, Italy has several world-class conservation and restoration institutes, including the International Centre for the Study of the Preservation and Restoration of Cultural Property in Rome and the Opificio delle Pietre Dure in Florence.

Yet political support for culture in Italy dwindled from the 1980s onwards, and funds continue to shrink alarmingly. Retiring staff working at cultural heritage sites are not replaced. The proportion of the state budget dedicated to culture shrank from 0.39% in 2000 (more than €2 billion, or US\$2.6 billion) to 0.19% in 2011 (less than €1.5 billion).

The consequences are evident at the Vesuvius archaeological sites. Any city will quickly deteriorate if its roofs are not fixed and its drains not cleared. Over the decades, water from below and above has caused salts to leach through walls, destabilizing them, damaging mosaics and destroying frescos.

The problems are as much managerial as financial. Pompeii acquired substantial subsidies through the European Union (EU) Structural Funds in the 1980s and 1990s. But instead of using those to conserve the exposed remains, the superintendency embarked on glamorous new excavation work to impress politicians. This went so badly that, at one point, the EU suspended payment.

In 1997, just 16 out of the Pompeii superintendency's 711 staff were archaeologists, architects and art historians; in the era of computers, 34 were typists. Successive governments went on to shamelessly ignore Pompeii's autonomy. The 2006 government siphoned off €30 million of Pompeii's income for spending elsewhere. In 2008, the government declared a one-year state of emergency for the site, later extended by a further year.

Responsibility for all aspects of cultural heritage in Italy is centralized within the ministry of culture, whose regional offices, called superintendencies, mediate local needs and prevent unauthorized activities.

This system protects heritage from crass development, but can be damagingly slow in operation. Moreover, staff at all sites — from

NATURE.COM
For a series on
little-known cultural
treasures, see:
go.nature.com/v52qt5

A. SOLARO/AFP/GETTY

archeologists to ticket collectors — are government employees with jobs for life. The inflexibilities make long-term planning almost impossible.

During the past decade or so, successive governments have experimented with new approaches to funding conservation, with some clear successes. The Egyptian Museum in Turin has, since 2005, been managed by a private foundation. This has renovated and modernized the museum, to international acclaim. And shoe magnate Diego Della Valle is paying €25 million for urgently needed conservation work on Rome's Colosseum that is being directed by the ministry. In return, he gets exclusive rights to use the image of the edifice to promote his products for 15 years. Alarmed academics have tried to equate such activities with privatization. But the heritage itself remains firmly in the possession of the state, which retains full power to control conservation or restoration projects.

Now the Pompeii superintendency has a further €105 million of EU structural funds to spend on securing its site, efficiently and effectively, under stern oversight — and within just three years. This will be a challenge, although the project acquired a further 20 or so architects and archaeologists this year.

Herculaneum, fortunately, won the support of philanthropist David W. Packard, son of the co-founder of the Hewlett-Packard information-technology company. His Packard Humanities Institute in Los Altos, California, has been running the Herculaneum Conservation Project in partnership with the superintendency and the British School at Rome for the past 11 years. This international, interdisciplinary team of archaeologists, architects and conservationists do unglamorous practical conservation work. This could be mending the ancient drainage networks, repairing roof coverings or driving out the pigeons whose voluminous, acidic excreta destroy frescos. The work is mostly low-tech — for example, the best solution they've found for the pigeons is to encourage falconers to visit the site regularly.

The Herculaneum project has inspired at least one other consortium of foreign scientists to bid to help to conserve and restore some frescoed houses in Pompeii, working in partnership with the Italians.

Such respectful international support for Italy's cultural heritage is fundamental. But the country will have to help itself by relaxing outdated labour laws and modernizing management of its cultural heritage systematically. Italy can't do much about Vesuvius' shadow. It can do a lot about the political shadows it casts on itself. ■ [SEE EDITORIAL P.302](#)

Alison Abbott is Nature's senior European correspondent.

Books in brief



Planet Without Apes

Craig B. Stanford HARVARD UNIV. PRESS 272 pp. \$25.95 (2012)
Will electronic gadgetry bring down the great apes? The link may seem surreal, but in this study of the plight of gorillas, chimpanzees, orangutans and bonobos, primatologist Craig Stanford reveals how mining coltan, a mineral used in electronics, destroys primate habitats and fuels the illegal bushmeat trade. In his wide-ranging call for action, Stanford — co-director of the Jane Goodall Research Center in Los Angeles, California — lays out the critical threats, arguing that humanity's closest cousins are viewed as savage 'others' and subjected to a genocidal urge last seen in the colonial era.



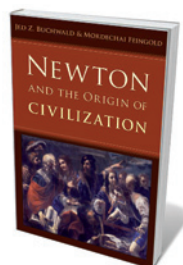
Jefferson's Shadow: The Story of His Science

Keith Thomson YALE UNIV. PRESS 322 pp. \$30 (2012)
Architect, philosopher, critic of slavery, slave-owner: the contradictions of American 'founding father' Thomas Jefferson are well known. That he was a scientist is not. Natural historian Keith Thomson redresses the balance in this finely wrought biography. Immersed in the work of Isaac Newton and Francis Bacon, Jefferson was arguably the most clued-up American naturalist of his time. This scintillating intellectual traced climate fluctuations, delighted in data tables, pored over fossils and helped to introduce the nation to palaeontology, geography, scientific archaeology and climatology.



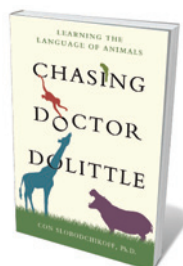
A Single Sky: How an International Community Forged the Science of Radio Astronomy

David P. D. Munns MIT PRESS 264 pp. \$34 (2012)
During the past 60 years, radio technology has transformed astronomy from a venerable practice reliant on visible light to an astounding new window on the cosmos. As historian David Munns reveals, it was all down to an international network of scientists who defied the rivalries of the cold war to ensure collaborative exploration of a 'single sky'. This remarkable science, forged by American, British, Australian and Dutch radio astronomers, ultimately led to the mapping of the Milky Way.



Newton and the Origin of Civilization

Jed Z. Buchwald and Mordechai Feingold PRINCETON UNIV. PRESS 544 pp. £34.95, \$49.50 (2012)
Isaac Newton spent most of his 84 years in pursuit of knowledge — mathematical to metaphysical. In this tome, historians Jed Buchwald and Mordechai Feingold unveil yet another strand: historical chronology. When Newton's *Chronology of Ancient Kingdoms Amended* was published in 1728, it drew fire for its dramatic revisions to timelines of civilizations past. Yet Newton, the authors show, approached the study — using astronomy and population dynamics — with the same rigour he brought to science.



Chasing Doctor Dolittle: Learning the Language of Animals

Con Slobodchikoff ST. MARTIN'S PRESS 320 pp. \$25.99 (2012)
An alarmed prairie dog can recognize and communicate the colour, shape, size and species of a predator. So says biologist Con Slobodchikoff, who — after 25 years of studying these hefty ground squirrels of the US grasslands — posits that animals have language. He bases his theory on a physiological and structural system not unlike the skeletal system that has parallels in humans and other vertebrates (think of human vocal chords and the avian double syrinx).



Potential effects of a tight orbit include volcanism caused by strong gravitational tides and fierce stellar winds, as imagined here for the exoplanet Gliese 876d.

SCIENCE FICTION

Curtains for space opera?

Since July, astronomers have killed off one trope of science fiction and given fresh life to another. **Leigh Phillips** gets *Mars Trilogy* author Kim Stanley Robinson's reaction.

Barnard's star is a star indeed. A member of the second closest star system to the Sun at 6 light years (1.84 parsecs) away, it pops up all over twentieth-century science fiction, from classic comics to Asimov.

Dutch-American astronomer Peter van de Kamp made the first modern claim to have spotted an exoplanet there in 1963, having studied the star since 1938. He thought that he had discovered wobbles in the position of Barnard that indicated a Jupiter-class planet in orbit around it. In 1969, van de Kamp revised his findings, positing two planets — one slightly bigger than Jupiter and one slightly smaller. But it wasn't long before other astronomers challenged the claims, suggesting that van de Kamp's 'discovery' was merely

an artefact of upgrade work at his observatory.

Since van de Kamp's time, the Barnard 'system' has been a staple of sci-fi, from short stories and novels, to films and television series. In Douglas Adams' *The Hitchhiker's Guide to the Galaxy* series and Arthur C. Clarke's *The Garden of Rama* (Bantam, 1991), it is a way station for interstellar travellers. Michael Moorcock uses an imagined planet orbiting the star as the site of a refugee camp for humans fleeing social breakdown on Earth. For Isaac Asimov, a Barnard-system planet is home to invertebrate marine animals. In a series of comic-book strips in

the 1970s, Will Eisner sited humankind's first contact with an extraterrestrial civilization on a planet in the system. And in the short-lived *Battlestar Galactica* spin-off series *Galactica 1980*, the dastardly Cylons are believed to be hiding there.

Recently, the status of this sci-fi staple itself wobbled. In August, a survey by a team of eight astronomers, led by Jieun Choi of the University of California, Berkeley, and covering 25 years' worth of measurements, concluded that Barnard's star does not have any planets — Earth-size or otherwise.

Two months later, astronomers had better news for the sci-fi cognoscenti. On 17 October, Xavier Dumusque at the University of Geneva in Switzerland and his team reported

➔ **NATURE.COM**
See *Nature's* science
fiction special at:
go.nature.com/mqc2jd

INGA NIELSEN

in *Nature* that Alpha Centauri B, a member of our closest star system, just 4.3 light years away, has an Earth-sized planet orbiting — albeit with a tight, sun-hugging ‘year’ of just 3.236 days, far from the presumed habitable zone (X. Dumusque *et al.* *Nature* **491**, 207–211; 2012).

This was sure to resonate with readers of Stanislaw Lem, Robert Silverberg, Philip K. Dick and, again, Asimov and Clarke, who all made use of the Alpha Centauri system in their fiction. It also appeared in the television series *Buck Rogers in the 25th Century*, *Doctor Who* and *Star Trek*. Indeed, Zefram Cochrane, the *Star Trek* character who ‘invented’ the warp drive, lived there.

So what do these two scientific developments mean for science fiction? Kim Stanley Robinson, author of the bestselling *Mars Trilogy*, takes a radical view. He suggests that we get over the idea of interstellar travel altogether: a probe would take 28,000 years to get to Alpha Centauri. “We can’t go fast enough to get to any of these places,” he says.

Barnard’s star was once “the place for nearby space”, Robinson says, as his novel *Icehege* (Ace, 1984) — in which characters build a starship headed for it — attests. Now that researchers have identified some 840 exoplanets, and NASA’s three-year-old Kepler space telescope has spotted 2,320 candidate planets, “there may never again be a single default destination”, Robinson continues.

In his recent book *2312*, which imagines humanity three centuries from now, spread across terraformed planets, asteroids and moons in our own Solar System, Robinson writes frankly about the galactic hinterland we inhabit. “The stars exist beyond human time, beyond human reach,” says the narrator. “We live in the little pearl of warmth surrounding our star; outside it lies a vastness beyond comprehension. The solar system is our one and only home.”

Of the idea that we are destined to go to the stars and inhabit, if not the whole Universe, maybe the whole galaxy, Robinson cautions “it’s a fantasy, of power, transcendence and a kind of species immortality. We have to get more realistic.” ■

Leigh Phillips is an International Development Research Centre fellow at *Nature*.
e-mail: leigh.phillips@gmail.com

COMPUTER SCIENCE

Virtually there

John Gilbey applauds a call for the digital to join the physical, biological and social in science.

In *On Computing*, Paul Rosenbloom examines the case for computing to enter the pantheon of great scientific domains alongside the physical, biological and social sciences. The centenary year of computing pioneer Alan Turing’s birth seems a fitting moment to put the idea to the test.

The study of computing, dated from Turing’s work, is only about 80 years old. It is variously claimed by engineering, physics, mathematics, linguistics and psychology — or seen merely as a supporting technology whose academic roots are irrelevant. Despite this, computing has arguably made more, and deeper, inroads into the daily life of humanity during the past 50 years than any other academic discipline, underlying a series of life-changing products. Imagine life today without mobile-phone networks, the Internet or medical imaging.

Drawing on his background in artificial intelligence, robotics and cognitive architecture, Rosenbloom leads us through the past, present and potential futures of computing as an academic discipline and demonstrates its linchpin position in a multidisciplinary environment.

He uses a novel ‘relational’ approach, unveiling the structures and connectedness across the various subfields of computing by looking at types of implementation and interaction within and between the existing major domains of science. To help clarify these relationships, Rosenbloom uses metascience expression language, a notation that facilitates the representation of the multidisciplinary fields and topics within science. Metascience expression offers both a technical context for Rosenbloom’s anecdotal material and a framework within which to debate the core tenets of the argument. Non-specialists who persevere with these sections of the book will benefit from a much more structured understanding of the make-up of the computing sciences.

Rosenbloom fields many examples of computing innovation — including immersive display technologies, neurally controlled prosthetics, and quasi-autonomous military systems such as advanced unmanned aerial vehicles, or

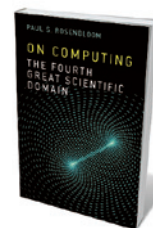
drones. These demonstrate that traditional demarcations between real and virtual environments will blur over the coming years as interfaces between human and machine are integrated to the point of invisibility. One example is the rapidly expanding field of augmented reality systems, early versions of which are already embedded in smartphones and tablets.

Rosenbloom’s reasoned analysis should help academia and the wider technical community to ensure that this transition is managed so as to deliver benefits to humanity in general. Otherwise, that enormous and life-changing power will be unfairly subjugated by a small minority of interests — technical, economic or political.

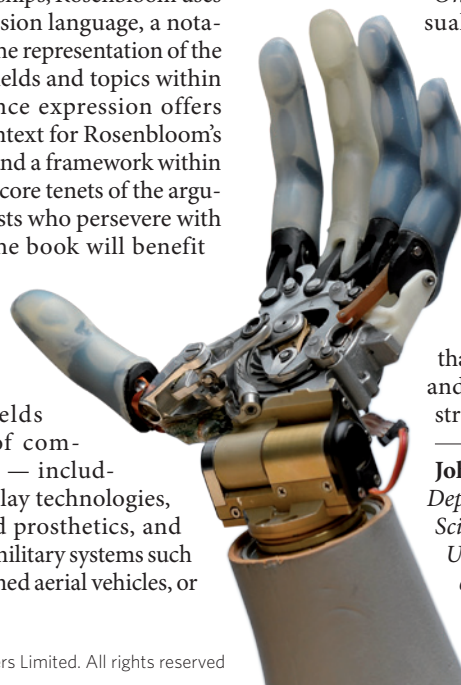
The text is permeated with a sense of delight in the opportunities offered by advances in the computing sciences. Rosenbloom offers elegant examples of the innovative ways in which computing developments and mature research areas can have hugely productive synergy — such as in surgical robotics and sophisticated prosthetic systems.

On Computing is an unusual, and welcome, mix of conventional academic text and personal odyssey. Any work citing Jane Austen and Richard Feynman in the same chapter easily passes my test for an interesting interdisciplinary read. Much more, this book offers an innovative set of tools that could kick-start debate and research on the future structure of the sciences. ■

John Gilbey teaches in the Department of Computer Science at Aberystwyth University, UK.
e-mail: gilbey@bcs.org.uk



On Computing — The Fourth Great Scientific Domain
PAUL S. ROSENBLUM
MIT Press: 2012.
312 pp. \$35, £24.95



E. SCHULZ/AP

Correspondence

Standardize the diet for zebrafish model

The standardization of diets for laboratory rodents in the 1970s minimized the contribution of unintended nutritional effects to experimental outcomes and made comparison between experiments more reliable (*Nature* **491**, 31–33; 2012). Despite success as a model species, zebrafish (*Danio rerio*) are still fed assorted commercially available diets of largely unknown nutrient composition. It is time to develop a standard formula diet for zebrafish in the laboratory, applying the extensive knowledge of fish nutrition from aquaculture.

We analysed the iron content of four commercial zebrafish diets and found that they contained 0.6–4.6 grams of iron per kilogram (g kg^{-1}) of dry feed. Because fish have a maximum iron requirement of 0.2 g kg^{-1} dry feed, these higher concentrations could be toxic. In salmon, for example, differences in dietary iron affect the cytochrome P450 detoxification system (A. Goksøyr *et al.* *Can. J. Fish. Aquat. Sci.* **51**, 315–320; 1994). One zebrafish diet was also deficient in vitamin C — a combination that would alter cellular redox status and could influence study parameters such as disease progression.

Failure to control for such variables compromises the validity of outcomes from zebrafish receiving different nutrition in an otherwise identical experiment.

Sam Penglase *National Institute of Nutrition and Seafood Research (NIFES), Bergen; and University of Bergen, Norway.*
spe@nifes.no

Mari Moren, Kristin Hamre
NIFES, Bergen, Norway.

Clean stoves already in use in rural India

The health and pollution problems caused by primitive heating stoves (*Nature* **490**, 343;

2012) are already being addressed in one rural Indian community.

In the state of Arunachal Pradesh in the eastern Himalayas — a biodiversity hot spot (N. Myers *et al.* *Nature* **403**, 853–858; 2000) — most people use biomass fuel as their primary source of energy. It is burnt in a safe, energy-efficient and smoke-free stove called a *chulha*.

This portable iron stove is enclosed, equipped with a heat-intensity control, an ash-collection tray and an exhaust pipe. It costs just 1,500–3,000 rupees (US\$28–56), and has a thermal efficiency of 60%, compared with 6–8% for traditional stoves. This translates into a significant saving of around 300 kilograms of wood fuel (biomass) equivalent per year (J. S. Rawat *et al.* *Curr. Sci.* **98**, 1554; 2010).

These improvements have proved to be a boon for rural women living in poor socio-economic conditions.

Sudhir Kumar *Jaypee University of Information Technology, Waknaghat, Himachal Pradesh, India.*
sudhir.syal@juit.ac.in

Union improves postdocs' rights

As president of the union UAW Local 5810, which represents more than 6,000 postdocs at the University of California, I recognize efforts by the US National Postdoctoral Association to improve our working conditions (*Nature* **489**, 461–463; 2012). But our union's experience has shown that recommending changes is not enough: organizations need formal negotiating power to make them effective.

UAW Local 5810 has bargained collectively for measurable improvements for postdocs. These include a contract with a minimum salary scale that matches that of the US National Institutes of Health's National Research

Service Award, a stable and comprehensive benefits plan, greater job security and the right to career-development resources.

The parental-leave policy negotiated by the union should help to address the under-representation of women in science and engineering and retain top talent. Also, our advocacy for an increase in federal research funding has earned the support of thousands of postdocs and more than 20 members of Congress from California.

These hard-won successes should be an encouragement to postdocs everywhere to organize union support (*Nature* **467**, 739–741; 2010).

Neal Sweeney *University of California, Santa Cruz, California, USA.*
president@uaw5810.org

Biomedical network in South America

The organization MERCOSUR — dubbed the Common Market of the South — promotes free trade and movement of goods, people and currency within a trade bloc of five countries in South America. The organization has now funded a large biomedical network spanning research institutes in Argentina, Brazil, Paraguay and Uruguay. We hope that this unprecedented initiative will encourage other regional scientific endeavours in South America.

The idea of the network is to help each other develop innovative biomedical projects that have potential for translational medicine. The network will encourage contributions from young investigators.

It aims to study the biological and epidemiological aspects of diseases that have social and economic impact; to create biotechnology platforms for clinical developments; and to build up human resources and

technology to a high standard.

In recognition of the importance of investment in science and technology on the development and welfare of communities, MERCOSUR will provide US\$7 million, with a further \$3 million coming from national funding. The MERCOSUR funding will come from its FOCES budget, better known for supporting local construction projects such as roads or hospitals.

Eduardo Arzt* *Biomedicine Research Institute of Buenos Aires, CONICET–Partner Institute of the Max Planck Society, Buenos Aires, Argentina.*
earzt@fbmc.fcen.uba.ar
*On behalf of 4 co-authors (see go.nature.com/s6ud4k for a full list).

Environmental stress seen since antiquity

I agree that biologists and sociologists need to get their acts together to determine the effects of environmental stress on our behaviour (*Nature* **490**, 143; 2012). You mention Francis Galton as the first to define its terms, but this discussion was going on long before his and Charles Darwin's time.

'Environmental determinism' has been an issue for at least 2,500 years. For example, it was discussed by Hippocrates and Strabo, by the Muslim historian Ibn Khaldun in his fourteenth-century book *Muqaddimah* and by the French philosopher Montesquieu in his political treatise *The Spirit of the Laws* in 1748.

Frank Vereecken *La Hulpe, Belgium.*
vereecken@yahoo.com

CONTRIBUTIONS

Correspondence may be sent to correspondence@nature.com after consulting the guidelines at <http://go.nature.com/cmchno>.

Edward Donnall Thomas

(1920–2012)

Immunologist who won Nobel prize for bone-marrow transplants.

Edward Donnall (Don) Thomas has been called the father of bone-marrow transplantation. Until the 1970s, every reported human marrow transplant had failed, and prominent immunologists declared that the barriers between individuals could never be crossed. Thomas persisted and eventually succeeded, sharing a Nobel prize for the feat in 1990. Since 1969, around one million patients with otherwise fatal blood disorders have received bone-marrow transplants.

Thomas died on 20 October 2012, aged 92, of heart failure. He was born in Mart, Texas, and his father was a general practice doctor, whom he often accompanied on house calls. Thomas received his bachelor's and master's degrees in organic chemistry from the University of Texas at Austin in 1941 and 1943, respectively. In 1942, he married fellow student Dorothy (Dottie) Martin. She helped to manage his research and papers throughout his career — the late George Santos of Johns Hopkins University School of Medicine in Baltimore, Maryland, once said: "If Dr Thomas is the father of bone-marrow transplantation, then Dottie Thomas is the mother." The couple had three children, two of whom are physicians.

In 1946, Thomas received his MD degree from Harvard Medical School in Boston, Massachusetts, followed by residency training at the city's Peter Bent Brigham Hospital and then service in the US Army. In 1950, he returned to the area as a research fellow at the Massachusetts Institute of Technology in Cambridge, and then as chief resident and instructor in medicine at Harvard.

In 1955, Thomas was appointed physician-in-chief at the Mary Imogene Bassett Hospital in Cooperstown, New York. Here he became fascinated by the discovery that rodents given a lethal dose of radiation could be rescued by an intravenous infusion of marrow cells from a donor. In 1957, Thomas treated a patient with leukaemia using high doses of total-body irradiation to wipe out the cancer, and then gave them an infusion of marrow cells from an identical twin. The transplant was at first successful, although the patient later died from a recurrence of the leukaemia.

Meanwhile, the medical literature was charting numerous cases of patients with blood disorders who had been treated using marrow transplantation from healthy family members. All the patients died from

infections or severe immune reactions that were not predicted from studies in inbred rodents. Many investigators left the field, pronouncing it a dead end.

Thomas did not give up. In 1957 he began experimenting in dogs. Like humans, dogs have unusual phenotypic



diversity, a well-mixed gene pool and can develop haematological diseases, including non-Hodgkin lymphoma. In late 1963, Thomas set up his laboratory at the United States Public Health Service (USPHS) Hospital in Seattle, which was affiliated to the University of Washington's medical school.

When I joined his small band of scientists in 1965 as a research fellow, transplantation was not a widely known concept. Indeed, the university's print shop once produced letterhead stationery for us that read "Division of Hematology and Transportation". A remarkable thing about Thomas's leadership style was that he was happy to give people like me a free hand in innovating, as long as it helped the patients.

Under Thomas's guidance, we spent the 1960s developing high-intensity irradiation treatments to eradicate patients' cancer cells and establishing the importance of tissue matching for transplant outcome. To control and treat graft-versus-host disease, we developed drug combinations to suppress the immune system and produced antibodies against human lymphocytes — which once

involved a 1.5-hour chase of an antibody-producing horse. Our total-body irradiation sources were set up in a Second World War underground bunker.

Work with patients began in 1969 at the USPHS hospital. Initial survival rates were low, and unexpected problems required going back and forth between bench and bedside — something that remained a hallmark of Thomas's work. By 1979, after performing a number of transplants, we were able to describe a phenomenon called graft-versus-tumour effects, in which donor lymphocytes help to eliminate residual malignant cells. There are patients who received transplants in 1971 still alive today.

Progress has been slow but steady. Early marrow donors were siblings. In 1979, the first patient with leukaemia was successfully treated with a transplant from an unrelated donor using new immunosuppressant drugs. This success sparked Thomas and others to set up international marrow-donor programmes. Transplant outcomes have improved, with survival rates for some diseases, such as aplastic anaemia, close to 95%.

In 1972, the US government closed the USPHS hospital. This prompted the founding of the private Fred Hutchinson Cancer Research Center (the Hutch), with close ties to the University of Washington's medical school. Thomas headed the medical oncology divisions at both.

Over the years, Thomas's team trained hundreds of young investigators. As one of them put it: "Virtually every major transplant centre in the world got its start by sending someone to train under Don Thomas." Owing to his influence, the Hutch's clinical focus has always been the patient, and its approach, one of teamwork. After retiring in 1989, Thomas continued writing manuscripts, lecturing and serving as an ambassador for the place.

Besides science, Don and Dottie had a passion for fishing and hunting, the fruits of which they shared at intimate dinners with colleagues and friends at their modest home. Reserved, hardworking and uncompromising, Thomas generously attributed much of his success to colleagues, nurses, support staff, patients and their families. ■

Rainer Storb is at the Fred Hutchinson Cancer Research Center and at the University of Washington, both in Seattle. e-mail: rstorb@fhcrc.org

COURTESY OF DOTTIE THOMAS

FORUM Evolutionary biology

Birds of a feather

A phylogenetic reconstruction of the diversification of birds across space and time provides a novel resource for evolutionary studies. But the methods used to construct this tree, and what insights can be inferred from it, are a source of debate. Two evolutionary biologists provide opinions on how to draw the lines. **SEE LETTER P.444**

THE PAPER IN BRIEF

- Geographical and ecological features, such as climate and niche opportunities, influence the evolutionary processes that generate new species.
- Jetz *et al.*¹ (page 444 of this issue) combined genetic and taxonomic information to construct a phylogenetic tree of the almost 10,000 species of extant birds that also considers

their historical locations*.

- The authors report that the diversification rates of bird species vary across the globe, with greater differences in rates between the Eastern and Western Hemispheres than across latitudinal lines.
- The phylogeny also reveals 'hot spots' of recent diversification in regions characterized by strong climatic fluctuation over the past 5 million years.

Disconnects in diversity

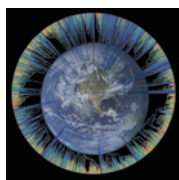
ROBERT E. RICKLEFS

Ever since biologists began cataloguing the diversity of life on Earth, they have sought to understand the origin and maintenance of global patterns of species richness — for example, that life is most diverse where the climate is warm and wet or where mountains vary the landscape. Biological diversity reflects a balance between the tendency of evolutionary lineages to form new species and the variety of living things that an environment can support. Furthermore, diversification depends on the availability of space, the dispersal of life forms among shifting continents, and variations in climate and resources at different locations and over time. Therefore, to understand the distribution of diversity is to interpret evolutionary diversification in historical and geographical contexts — and this is the key to Jetz and colleagues' remarkable accomplishment.

By integrating their reconstruction of the ancestral relationships of all of the approximately 10,000 known extant species of bird with maps of their distributions, Jetz *et al.* derived a detailed picture of average diversification rates over the surface of the globe. The patterns revealed are intriguing. One might expect more species where the diversification rate is higher, but the authors find that

diversity and diversification rate correspond poorly around the planet — evolutionary lineages split more frequently, on average, in the Western than the Eastern Hemisphere, but not in the tropics compared with higher latitudes. They also find that the overall diversification rate is higher in passerines (songbirds) than in non-passerines (ducks, raptors, shorebirds and others), as expected given the species richness of the former group. But, surprisingly, the data show that the relative contribution of each group to the diversification rate differs between regions. Particularly intriguing is the relative evolutionary quiescence of modern passerines in Australia and New Guinea, where the passerines, now the largest group of birds, originated around 60 million years ago² (Fig. 1).

Biologists will debate whether Jetz and colleagues' phylogenetic reconstruction is up to the task. Although some DNA-sequence information is available for around two-thirds of bird species, the genomes of only a few have been well sampled. Jetz *et al.* built their phylogeny on a backbone of 158 major bird clades whose relationships had previously been defined³ (a clade represents a 'branch' of a phylogenetic tree, including an ancestor and all its descendants). To this, they attached new phylogenetic detail, using a complex algorithm based on taxonomic distinctions to determine the placement of species for



NATURE.COM

For more on bird diversification, see: go.nature.com/jvlqnv

which sequence data are not available.

The result is perhaps not perfect, but it is probably the best possible for now, and is certainly the most ambitious. Moreover, the sequence data that would be required for a substantial improvement might not be worth the effort, because any inference on rates of species production depends on how we circumscribe species. For example, the authors' method for calculating diversification rate — on the basis of the lengths of branches close to the tips of the phylogeny — may provide inflated estimates in regions in which bird populations are finely distinguished at the species level, such as Europe and North America. So additional sequencing without taxonomic revision might be an empty exercise.

"The result is perhaps not perfect, but it is probably the best possible for now."

The apparent absence of a latitudinal gradient in the diversification rates of birds is consistent with a recent analysis for mammals⁴. This result implies that the great species richness of tropical environments is a matter of age: Earth was mostly tropical before temperate and boreal environments began to expand around 30 million years ago. However, an assessment based on a different phylogenetic approach⁵ concluded not only that the recent speciation and extinction rates of birds both increase towards higher latitudes, but also that the difference between the two (the diversification rate) has nevertheless been higher in the tropics. These conflicting interpretations and other enigmatic patterns arising from Jetz and colleagues' new phylogeny — for example, that extinction compared with speciation has been relatively infrequent, or that the overall rate of diversification has increased towards the present, particularly with the expansion of cold and arid environments — will motivate further work for some time to come.

Robert E. Ricklefs is in the Department of Biology, University of Missouri-St. Louis, St. Louis, Missouri 63121, USA. e-mail: ricklefs@umsl.edu

*This article and the paper under discussion¹ were published online on 31 October 2012.



Figure 1 | Birds in space and time. These Cape sparrows (*Passer melanurus*) are one of around 5,000 species of songbird (passerine), which make up around half of the known species of bird. Jetz and colleagues¹ have constructed a phylogenetic

tree of all known avian species and mapped this to spatial data of species distribution to assess how bird diversity and diversification rates compare at different times in history and in different regions.

First steps for birds

MARK PAGEL

Jetz and colleagues' bird phylogeny joins several other attempts to reconstruct the history of entire classes of organism, including the 5,000 or so mammals⁶ and the roughly 6,000 amphibian species⁷. These large phylogenetic trees have a value that extends beyond describing the evolutionary relationships among a group of organisms: they grant unprecedented statistical power to attempts to reconstruct the probable historical events and processes of evolution⁸, such as our understanding of ancestral states, or rates of morphological change and speciation. This statistical power allows researchers greater confidence in ruling some proposals in and others out. But it also means that, if the phylogeny is wrong, it might confidently return wrong, biased or misleading answers to tests of evolutionary questions.

And this is why we must receive Jetz and colleagues' tree with a measure of caution. Rather than seeking to infer the avian tree from gene-sequence or other information, such as data on morphology and behaviours, Jetz *et al.* relied on the findings of previous studies to fix the tree's major outlines and the broad placement of most of its species. In fact, the authors assigned positions in the tree to roughly one-third of the bird species on the basis of previous taxonomic classifications alone. As a consequence, the authors' method never explores the possible universe of avian relatedness, and so we are left wondering whether there might be other trees that provide equally good or even better descriptions

of avian evolution, given what we know about birds' genetic, morphological and behavioural similarities and differences.

The authors could have avoided this nagging worry by gathering data on as many species as possible and then inferring the tree by a more conventional route that did not place such strong prior constraints on the outcome. Curiously, for instance, Jetz *et al.* seem to have ignored recently published gene-sequence data⁹ on around 4,000 passerine bird species that were used to infer a tree of this group — the largest within the class Aves.

Jetz and colleagues press their tree into service to study avian speciation rates, concluding that these rates have increased through time

"If the phylogeny is wrong, it might confidently return biased or misleading answers to tests of evolutionary questions."

and that they are, for example, higher in the Western than the Eastern hemisphere, with latitude generally having a smaller influence. These are the sorts of broad questions that such large trees should be used to test, but here we must not lose sight of the statistical power this tree grants. For example, the authors' methods for estimating speciation rates depend on how well one can estimate the lengths of the branches of the phylogeny in units of time. As has been previously discussed¹⁰, there are reasons to be cautious about the branch lengths that are returned by methods used to infer time-dated trees, and even small biases in branch-length estimation, when integrated over so many species, can produce apparent trends that may or may not be real.

So it is difficult to know what to make, in evolutionary terms, of the 'hemisphere' effect

the authors report. It is not just that there are questions about the tree and the estimation of speciation rates — hemisphere boundaries are arbitrary constructions, and the regions they separate, whether north–south or east–west, are vast and each harbour a wealth of ecological and climatic conditions. Furthermore, when seeking associations between characteristics of species, or between characteristics of species and their environments, it is vital to identify multiple evolutionarily independent instances of the two traits changing in tandem¹¹. But Jetz *et al.* have not done this for the features of climate and ecology that they suggest influence speciation rates.

So these are still 'first steps' towards a phylogeny of birds and our understanding of their rates of speciation. The important questions Jetz and colleagues raise invite careful second steps towards confirming or refuting their proposals in this tricky area. ■

Mark Pagel is in the School of Biological Sciences, University of Reading, Reading RG6 6AS, UK.
e-mail: m.pagel@reading.ac.uk

1. Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K. & Mooers, A. O. *Nature* **491**, 444–448 (2012).
2. Ericson, P. G. P., Irestedt, M. & Johansson, U. S. *J. Avian Biol.* **34**, 3–15 (2003).
3. Hackett, S. J. *et al.* *Science* **320**, 1763–1768 (2008).
4. Soria-Carrasco, V. & Castresana, J. *Proc. R. Soc. B* **279**, 4148–4155 (2012).
5. Weir, J. T. & Schluter, D. *Science* **315**, 1574–1576 (2007).
6. Bininda-Emonds, O. R. P. *et al.* *Nature* **446**, 507–512 (2007).
7. Fritz, S. A. & Rahbek, C. *J. Biogeogr.* **39**, 1373–1382 (2012).
8. Pagel, M. *Nature* **401**, 877–884 (1999).
9. Huggall, A. F. & Stuart-Fox, D. *Nature* **485**, 631–634 (2012).
10. Venditti, C., Meade, A. & Pagel, M. *Nature* **463**, 349–352 (2010).
11. Harvey, P. M. & Pagel, M. D. *The Comparative Method in Evolutionary Biology* (Oxford Univ. Press, 1991).

CLIMATE SCIENCE

Historical drought trends revisited

A new assessment of drought trends over the past 60 years finds little evidence of an expansion of the area affected by droughts, contradicting several previous estimates. [SEE LETTER P.435](#)

SONIA I. SENEVIRATNE

The 2011–12 drought in the United States was reported as one of the most severe on record^{1,2}, and led to economic losses of billions of dollars^{3,4}. When such extreme events occur (Fig. 1), a common question is whether they might be a result of climate change¹. Not only is this question far from trivial⁵, but it is also more complex for droughts than for most other climate extremes. On page 435 of this issue, Sheffield *et al.*⁶ report an overestimation of historical drought trends obtained using a method that served as the basis for historical drought assessments made in the Fourth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC)^{7,8}, as well as for analyses in more recent publications^{9,10}.

The authors' results suggest a high uncertainty in global-scale drought trends over the past 60 years and little evidence of an increase in the total area affected by droughts. A recent IPCC assessment⁵ highlighted strong uncertainties in historical drought records and, in particular, inconsistent conclusions of studies carried out after the Fourth Assessment Report with respect to some regional drought trends. Sheffield and colleagues' results are consistent with this assessment, but also suggest that methodological issues may partly explain reported conflicting results in the literature.

Soil-moisture drought, which is of most

relevance to agriculture, is induced by a deficit in the land water balance and is caused by lack of precipitation and/or excess evapotranspiration⁵ (Fig. 2). Evapotranspiration refers to the moisture loss from soils, either through plant transpiration (water extracted by the plants and lost through the leaves' stomata) or by direct evaporation from moist surfaces (such as bare soils, lakes, rivers, or water stored on top of leaves).

The key point addressed by Sheffield and colleagues is the contribution of evapotranspiration as a driver for droughts. Most approaches compute actual evapotranspiration as a function of potential evaporation — that is, the evaporation occurring from bodies of water. In their study, the authors evaluate the extent to which different formulations used for computing potential evaporation can affect resulting estimates of historical drought trends that are computed by means of a drought indicator called the Palmer Drought Severity Index (PDSI). It should be noted that evapotranspiration depends not only on potential evaporation in the PDSI but also on soil-moisture availability¹¹, an aspect not directly assessed by the authors.

The PDSI approach is commonly used to assess drought trends^{7–11}, although its validity for such applications is questionable because of several issues^{5–7,12}. One of these is the usual practice of estimating potential evaporation as if it were solely dependent on temperature

and latitude — an estimation known as the Thornthwaite-based formulation. This formulation neglects the role of several other climate drivers of potential evaporation, in particular wind speed, relative humidity and solar and long-wave radiation (Fig. 2). In addition, using temperature as a driver for drought overlooks the fact that, in dry conditions, the causal link is often reversed — that is, drought itself induces hot temperatures when the lack of soil moisture leads to a suppression of evaporative cooling¹³.

In earlier studies, the rationale for using a formulation that relied only on temperature was the absence of credible long-term global data sets of other driving variables⁷. Sheffield *et al.* used a recently compiled data set for these variables to assess their impact on trends in potential evaporation and resulting drought computed using the PDSI. By applying the Penman–Monteith formulation, which considers these other drivers, the authors find much weaker trends than with the Thornthwaite formulation, and little evidence of an expansion of the area affected by drought in past decades^{7–10}.

The authors provide detailed evaluations of their results, as well as thorough explanations for the apparent conflict with previous studies^{9,11} that found few differences between PDSI drought trends obtained with the Penman–Monteith and Thornthwaite formulations (see the Supplementary Information to the paper⁶). On the basis of this analysis, the main discrepancies in these previous studies^{9,11} seem to be rooted in slight differences in methodological approach, for example differences in the calibration periods used or the atmospheric forcing (data-set choice; data-gap filling, in some cases with climatological data; and consideration of spurious trends in data sets). Clearly, the quality and reliability of the forcing data sets⁶ and lack of observations⁵ remain an issue for any assessment related to drought trends. Because of these uncertainties, Sheffield and colleagues' investigation has its own limitations, as the authors themselves recognize. Therefore, their results will also need to be confirmed by other independent analyses.

The 2007 IPCC Fourth Assessment Report's conclusion⁸ that the area affected by droughts was "likely" to have increased in many regions since the 1970s had already been revised in a more recent IPCC special report on extreme events and disasters published earlier this year (the SREX Report)⁵. This report assessed that "[t]here are still large uncertainties regarding observed global-scale trends in droughts", and highlighted regions in which drought trends have increased (southern Europe and West Africa) as well as those showing decreasing trends (central North America and north-western Australia)⁵. These reported regional trends agree with Sheffield and colleagues' results. Furthermore, the SREX Report did not provide any assessment of previous changes



Figure 1 | Severe drought. A field of dried maize (corn) plants near Percival, Iowa, during the severe drought that struck the United States in 2012.

N. HARNIK/AP

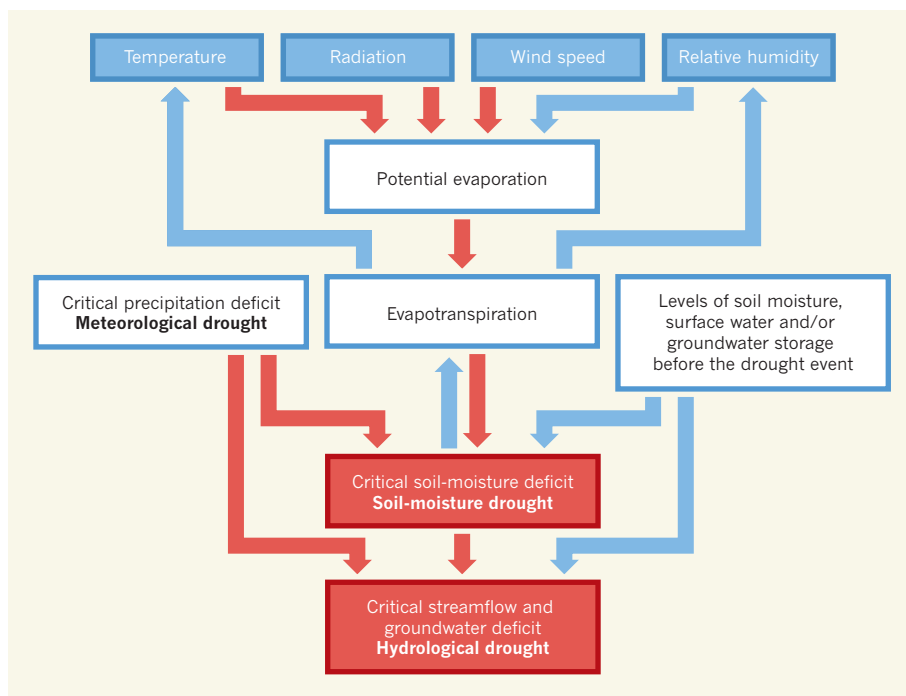


Figure 2 | Drought drivers. Several factors, apart from temperature, affect the development of droughts. These should all be taken into account for assessing historical drought trends and the contribution of anthropogenic climate change to recent drought events. Red arrows indicate factors that contribute to drought, and blue arrows show factors that counteract it. Sheffield *et al.*⁶ find that results obtained for historical trends in the global land area affected by drought (even possibly affecting the sign of these trends) are heavily influenced by the meteorological drivers chosen to estimate them. (Adapted from ref. 5.)

in the areas affected by droughts. Given Sheffield and colleagues' findings, this metric of drought (that is, the total land area affected by drought) seems rather ill defined, because the error range of their Penman–Monteith-based PDSI estimates does not exclude either positive or negative trends in this quantity.

The authors' results confirm the complexity of the processes that lead to changes in drought conditions, also discussed in the SREX Report⁵. The findings imply that there is no necessary correlation between temperature changes and long-term drought variations, which should warn us against using any simplifications regarding their relationship. Furthermore, apart from the variables considered in potential evaporation, the PDSI has several other shortcomings, in particular those resulting from the simplicity of its water-balance model^{5,12}. These may explain a reported tendency of the PDSI approach (even when using the Penman–Monteith formulation¹²) to overestimate future drying trends when driven with climate-model output, compared with other estimates, including soil-moisture output of the climate models themselves. Future investigations should carefully consider these uncertainties, for instance in the context of palaeoclimate studies⁶ and when relating specific changes in global mean temperature (for example, for the commonly discussed 2 °C target¹⁴) to their effects on drought. ■

Sonia I. Seneviratne is at the Institute for

Atmospheric and Climate Science, ETH Zurich, 8092 Zurich, Switzerland.
e-mail: sonia.seneviratne@env.ethz.ch

1. Rupp, D. E. *et al.* in *Explaining Extreme Events of 2011 From a Climate Perspective* (eds Peterson, T. C., Stott, P. A. & Herring, S.) *Bull. Am. Meteorol. Soc.* **93**, 1052–1054 (2012).
2. www.ncdc.noaa.gov/sotc/drought/2012/7.
3. <http://today.agrilife.org/2012/03/21/updated-2011-texas-agricultural-drought-losses-total-7-62-billion>.
4. www.economist.com/node/21559381.
5. Seneviratne, S. I. *et al.* in *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change* (eds Field, C. B. *et al.*) 109–230 (Cambridge Univ. Press, 2012).
6. Sheffield, J., Wood, E. F. & Roderick, M. L. *Nature* **491**, 435–438 (2012).
7. Trenberth, K. E. *et al.* in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Solomon, S. *et al.*) 235–336 (Cambridge Univ. Press, 2007).
8. Solomon, S. *et al.* 'Technical Summary' in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Solomon, S. *et al.*) (Cambridge Univ. Press, 2007).
9. Dai, A. *Wiley Interdisc. Rev. Clim. Change* **2**, 45–65 (2011).
10. Dai, A. *Nature Clim. Change* <http://dx.doi.org/10.1038/NCLIMATE1633> (2012).
11. van der Schrier, G., Jones, P. D. & Briffa, K. R. *J. Geophys. Res.* **116**, D03106 (2011).
12. Burke, E. J. *J. Hydrometeorol.* **12**, 1378–1394 (2011).
13. Mueller, B. & Seneviratne, S. I. *Proc. Natl Acad. Sci. USA* **109**, 12398–12403 (2012).
14. Meinshausen, M. *et al. Nature* **458**, 1158–1162 (2009).

Spontaneous flows and self-propelled drops

The construction of *in vitro* assemblies of biological components that exhibit properties of living matter may shed light on the physical aspects of the dynamic reorganization that continuously occurs inside cells. [SEE LETTER P.431](#)

M. CRISTINA MARCHETTI

Liquids and gels can be made to flow by applying external forces at their boundaries. In this issue, Sanchez *et al.*¹ report the observation of self-sustained flows that occur in the absence of external forces — a hallmark of living systems — in a model gel. When this ‘active’ gel is confined to the interior of water droplets in a water–oil emulsion, the flows resemble the streaming used by cells to circulate their fluid content. Even more remarkable is the fact that, when one of these gel-filled droplets comes into contact with a hard surface, the self-driven flows of the confined gel drive the droplet along the surface*.

To build their gel, Sanchez *et al.* sequentially assembled ingredients extracted from cells (Fig. 1). The first — and key — components are microtubules. These stiff, cylindrical filaments are one of the constituents of the cytoskeleton, the polymer network that mediates force transmission and motility in cells. Microtubule dynamics in cells is regulated by several proteins. Among these is kinesin, a motor protein capable of ‘walking’ on individual microtubules by converting chemical energy from ATP fuel molecules into

mechanical work. To construct the active units of their gel, the authors used a protein called streptavidin as a scaffold to assemble clusters of kinesins that could simultaneously bind to multiple microtubules.

Finally, Sanchez *et al.* added nanometre-sized polymer coils to the solution. This step was essential to promote the formation of microtubule bundles that, in the presence of ATP, are continuously remodelled by the action of the crosslinking motor proteins. The coils induce attractive forces between the microtubules through a mechanism known as depletion interaction. This interaction arises when two filaments come near to each other, because the narrow gap between them is no longer accessible to the polymer coils. This creates an osmotic pressure difference that effectively acts as an attractive force between the filaments². Sanchez and colleagues’ overall hierarchical assembly process was recently used by the same group to build artificial cilia that beat periodically and, when densely packed on a substrate, spontaneously synchronize their beating pattern to create travelling waves³.

At a moderate density, the microtubule bundles form a polymer network that is

internally driven by the action of kinesins. The network flows spontaneously and exhibits mixing and enhanced transport, compared with its non-active counterpart (which is obtained when the ATP fuel runs out); this was demonstrated by the authors by tracking small particles suspended in the gel.

At scales much larger than the typical bundle length (tens of micrometres), the rich dynamics of the system resembles that of complex fluids such as liquid crystals driven by externally applied fields, but differs from them in that it occurs spontaneously as a result of the internal drive. This is the key property of active materials that are driven out of equilibrium not by forces applied at their boundaries, but rather by an input of energy on each unit, as in a suspension of swimming bacteria. Energy uptake at the microscopic scale is crucial for driving emergent phenomena and self-organization in disparate systems⁴ — from naturally occurring ones, such as bacterial suspensions and flocks of birds, to chemical and mechanical analogues, such as self-propelled Janus colloids (microscopic particles that have two faces with distinct properties).

When Sanchez *et al.* confined the microtubule network to a water–oil interface, the resulting dense, two-dimensional film again exhibited self-sustained streaming flows that seemed to be associated with bundle fracturing and healing. The complex dynamics yielded patterns resembling topological defects — structures that can be generated in liquid crystals at equilibrium by confinement or external drive.

Finally, when the researchers confined the active gel to droplets of at least 30 micrometres in diameter, the gel was spontaneously adsorbed to the inner surface of the droplets, turning into a two-dimensional active film on a curved substrate. Remarkably, the self-sustained active flows of the trapped gel drove autonomous movement of the droplet on a substrate. Although the motile droplets moved along somewhat circular trajectories, rather than travelling in a straight line, they covered about 250 micrometres in 33 minutes. These moving drops bring to mind recent theoretical work⁵ showing that active drops in a fluid spontaneously acquire directed motility. For drops in which the active constituents — the microtubule bundles in Sanchez and colleagues’ work — form large domains and have, on average, a common orientation but no preferred direction, the theory indeed predicts rotational motion of the drops.

Reconstituted microtubule–kinesin systems have been explored before as models for active self-assembly, not least in the remarkable experiments^{6,7} that led the way to current studies of pattern formation in active systems. In those experiments, kinesin complexes driven by ATP organized microtubules into

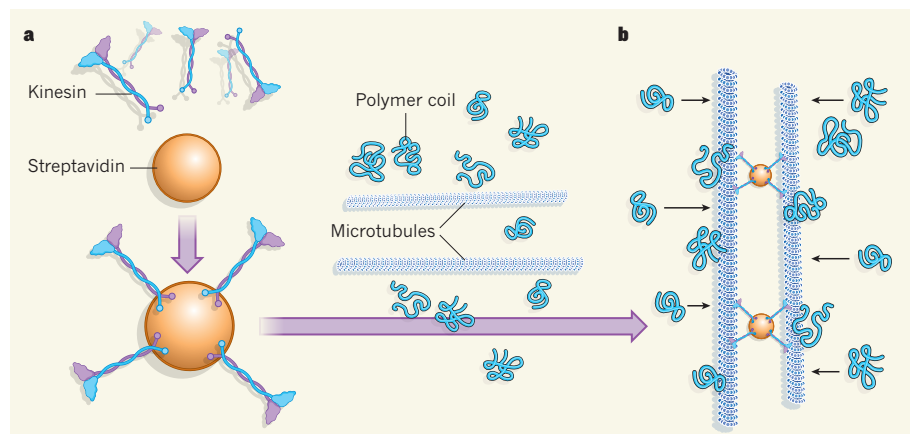


Figure 1 | Assembly of microtubule bundles. **a**, Sanchez *et al.*¹ combined the protein streptavidin with kinesin motor proteins that had been modified to bind to streptavidin (modification not shown). The proteins self-assembled to form clusters of several kinesin molecules in complex with streptavidin. **b**, The authors then added microtubule filaments and polymer coils to the mix. The polymer coils generated ‘depletion’ forces that pushed the microtubules together, promoting the formation of microtubule bundles mediated by the kinesin clusters. The bundles formed the basis of an ‘active’ gel — a material that generated self-sustained, internal flows of fluid in the absence of external forces.

*This article and the paper under discussion¹ were published online on 7 November 2012.

spirals and asters reminiscent of a cell's mitotic spindle, a star-like microtubule assembly that mediates cell division. One important difference is that the structures seen in the earlier work^{6,7} were essentially static, whereas Sanchez and colleagues' microtubule gel generates continuously evolving, spontaneous flows that persist as long as ATP is present — not unlike what happens in living cells. Furthermore, Sanchez *et al.* report that the internally generated flows in their active gel can be tuned by varying the ATP concentration, confirming the self-sustained, non-equilibrium nature of the dynamics. The fact that microtubules are assembled into bundles seems to be essential for yielding self-sustained motion (see

NATURE.COM
For more on active
gels, see:
go.nature.com/ougi8

Movie S2 in the Supplementary Information to the paper¹), but the reason for this remains an open question. Also

unexplained is why the behaviour of the active microtubule network is so different from that of gels composed of actin filaments and myosin motor proteins, in which activity yields spontaneous contraction⁸.

Sanchez and colleagues' work is a beautiful example of a growing class of experiment in biomimetic assembly, aimed at building systems that exhibit some of the features of living matter. Will it be possible to control and direct the motility of the active droplets? And can the flow-induced structures be harnessed and used as guides for the transport of particles through fluid, as those in cells are? This remains to be seen. Meanwhile, experiments of this type are beginning to shed light on the physical aspects of the complex dynamical reorganization that occurs continuously inside cells. When combined with studies of the biochemical machinery and signalling that drive such reorganization, they may ultimately

lead to a quantitative understanding of the mechanics of living matter. ■

M. Cristina Marchetti is in the Department of Physics, Syracuse University, Syracuse, New York 13244-1200, USA.
e-mail: mcm@phy.syr.edu

1. Sanchez, T., Chen, D. T. N., DeCamp, S. J., Heymann, M. & Dogic, Z. *Nature* **491**, 431–434 (2012).
2. Lekkerkerker, H. N. W., Poon, W. C.-K., Pusey, P. N., Stroobants, A. & Warren, P. B. *Europhys. Lett.* **20**, 559–564 (1992).
3. Sanchez, T., Welch, D., Nicastro, D. & Dogic, Z. *Science* **333**, 456–459 (2012).
4. Marchetti, M. C. *et al.* Preprint at <http://arxiv.org/abs/1207.2929> (2012).
5. Tjhung, E., Marenduzzo, D. & Cates, M. E. *Proc. Natl Acad. Sci. USA* **109**, 12381–12386 (2012).
6. Nédélec, F. J., Surrey, T., Maggs, A. C. & Leibler, S. *Nature* **389**, 305–308 (1997).
7. Surrey, T., Nédélec, F., Leibler, S. & Karsenti, E. *Science* **292**, 1167–1171 (2001).
8. Kasza, K. E. & Zallen, J. A. *Curr. Opin. Cell Biol.* **23**, 30–38 (2011).

ANIMAL BEHAVIOUR

Personality in the wild

Behavioural traits can influence an individual animal's fitness, and trait combinations can change over its lifetime, according to a study of wild trout during a key period in their development.

ALISON M. BELL

Life is hard for a young brown trout in a cold Swedish stream. There are so many dangers to watch out for, such as a hungry mink lurking around the bend, and so many things to do, such as competing for food. Indeed, a young brown trout has only around a 10% chance of surviving to adulthood¹. If a fish can beat the odds and survive this dangerous period, it emerges different from before — not just bigger, but also behaviourally changed. A recent paper by Adriaenssens and Johnsson in *Ecology Letters*² reports that the individuals that make it through this bottleneck behave more predictably across contexts than they did before. Is this because of those harrowing early experiences? Or are these fish the ones that were better adapted in the first place? Adriaenssens and Johnsson's findings suggest that the answer is an intriguing combination of both factors.

Animal personalities are interesting to researchers because behaviour is notoriously flexible — unlike most morphological traits, behaviour can change almost instantaneously. Within seconds, a fish might go from aggressively attacking an intruder to foraging alone in the middle of the stream. But there is growing evidence that behaviour does not always change

at a moment's notice, and that animals have distinctive personalities that they retain over time. One view³ is that animal personalities may result from constraints: limiting mechanisms that prevent an individual from being able to change, such as a genetic propensity. An alternative interpretation⁴ is that consistent differences in behaviour between individuals might be the result of adaptation through natural selection. There is evidence for both the constraint⁵ and the adaptive⁶ models,



Figure 1 | Fishy activity. Adriaenssens and Johnsson's study² of the behaviour of wild brown trout (*Salmo trutta*) shows that individuals with consistently high activity levels are more likely to survive the early months of life.

and strong support for the latter comes from research showing that behavioural consistency arises from both behavioural plasticity (when individuals change their behaviour in response to environmental conditions) and non-random survival of individuals⁷. However, that study was carried out in the lab, where life is relatively simple. The significance of Adriaenssens and Johnsson's work is that it starts to show us how adaptive personalities can emerge in the wild.

The authors captured young (around two and half months old) brown trout (*Salmo trutta*; Fig. 1) in a stream in western Sweden and gave each individual a unique colour mark. The trout were then put through a series of behavioural assays in the lab. One of these was an 'open-field test', in which trout were individually placed in an open arena and observed to determine whether they were the kind of fish that explores everything, or the type that moves little and hunkers down in one spot. Another assay involved a confrontation with an opponent — in this case, the trout's own reflection in a mirror. Here, the researchers were looking to see whether the individual attacked the intruder or if it was relatively non-aggressive. After assessing each fish in all of the assays, the researchers released them back into the stream.

Two months later, Adriaenssens and Johnsson returned to the stream. Of the 81 individuals that were tested, they recaptured 28. On the basis of the assumption that those fish that were not recaptured had died, the authors' analyses showed that an individual's behaviour predicted its survival: trout that had been very active in the open-field test were more likely to survive to 4.5 months of age than those that had moved around less. An alternative explanation would be that the inactive individuals did not die, but rather were

harder to recapture or more likely to disperse out of the study area. However, Adriaenssens and Johnsson provide evidence against both of these possibilities, showing that inactive fish were in fact easier to catch and did not move as far in the stream. So the first notable result from this study is that it shows natural selection acting on differences in behaviour among individuals in a wild population.

The authors then put the recaptured fish through the same assays and found that their behaviour had changed during their time back in the stream — the trout were more active at 4.5 months of age than they had been at 2.5 months. However, despite this overall behavioural change and the vagaries of life in the wild, the survivors retained their relative personality traits: the fish that had been the most active in the first round of testing, for example, were still the most active.

Finally, and perhaps most intriguingly, Adriaenssens and Johnsson report that the individuals' behaviour became more distinctive. Whereas the young trout did not behave consistently across the different behavioural assays, the older trout did, such that highly

exploratory individuals were now also more aggressive. In other words, the individuals were now more predictable, and a 'behavioural syndrome'³ linking exploratory behaviour to aggressive behaviour had emerged.

There are two ways in which this could have occurred. First, it may be that only those individuals that were relatively exploratory and aggressive, or relatively non-exploratory and non-aggressive, in the first place survived. Alternatively, it is possible that individuals changed their behaviour over time and this caused different behaviours to become coupled together. Adriaenssens and Johnsson present evidence in support of both mechanisms, but they were unable to thoroughly disentangle the two; designing experiments that can tease these two processes apart is a pressing goal for future work. Further studies are also needed to determine whether there are consequences of such non-random survival in the next generation — in other words, if behavioural variation is heritable.

The biggest question, though, is why did individual fish become predictable? That is, why did behaviours become coupled together,

and what are the fitness advantages to being a trout that behaves consistently? Several hypotheses exist to explain why behaviours should come packaged together⁸ — for example, that there might be social benefits of being predictable⁹. But so far there have been few empirical tests of these ideas, and this represents another exciting challenge for future work. ■

Alison M. Bell is in the School of Integrative Biology, University of Illinois Urbana-Champaign, Urbana, Illinois 61801, USA. e-mail: alisonmb@life.illinois.edu

1. Elliott, J. M. *Quantitative Ecology and the Brown Trout* (Oxford Univ. Press, 1994).
2. Adriaenssens, B. & Johnsson, J. I. *Ecol. Lett.* <http://dx.doi.org/10.1111/ele.12011> (2012).
3. Sih, A., Bell, A. M., Johnson, J. C. & Ziemba, R. E. *Q. Rev. Biol.* **79**, 241–277 (2004).
4. Bell, A. M. *J. Evol. Biol.* **18**, 464–473 (2005).
5. Pruitt, J. N. *et al. J. Evol. Biol.* **23**, 748–756 (2010).
6. Dingemanse, N. J. *et al. J. Anim. Ecol.* **76**, 1128–1138 (2007).
7. Bell, A. M. & Sih, A. *Ecol. Lett.* **10**, 828–834 (2007).
8. Dingemanse, N. J. & Wolf, M. *Phil. Trans. R. Soc.* **365**, 3947–3958 (2010).
9. D-all, S. R. X., Houston, A. I. & McNamara, J. M. *Ecol. Lett.* **7**, 734–739 (2004).

the skin by absorbing ultraviolet (UV) rays⁸, the skin of red-haired people has a particular tendency to accumulate light-induced damage.

Mitra *et al.* show that completely avoiding UV rays would not protect red-haired people from melanoma. As an animal model of red-haired individuals, the authors used mice with mutated, non-functional *Mclr*, and thus with melanocytes that cannot induce eumelanin synthesis. Whereas control mice had a dark-brown, nearly black coat colour, the *Mclr*-mutant mice had golden-yellow coats. A third group of mice expressed non-functional tyrosinase — an enzyme essential for both eumelanin and pheomelanin synthesis — and had white coats, mimicking albinism in humans. These 'white' mice and the 'black' mice possessed intact *Mclr*.

Mitra *et al.* crossed their differently coloured mice with mice that expressed *BRAF*^{V600E} — one of the most common gene mutations in melanoma — in their melanocytes⁹. *BRAF*^{V600E} is carried by 40–60% of patients with this cancer¹⁰. Furthermore, patients with melanoma who carry RHC variants also have a high frequency of *BRAF* mutations¹¹, suggesting that *BRAF* mutations have a role in the development of melanoma in the red-haired population.

When *BRAF*^{V600E} expression was induced in melanocytes, the black mice and their white counterparts developed melanoma only at low rates and after long periods of time. By contrast, more than 50% of the redhead mice developed this cancer within a year of *BRAF*^{V600E} induction, despite being kept in a UV-free environment. Surprisingly, blocking pheomelanin synthesis not only gave

CANCER

Complexion matters

Sun exposure indisputably increases the risk of skin cancer. Mouse studies suggest that, in red-haired individuals, genetic factors also contribute through a mechanism that acts independently of exposure to sunlight. SEE LETTER P.449

**MIZUHO FUKUNAGA-KALABIS
& MEENHARD HERLYN**

The most common environmental risk factor for skin cancer is overexposure to sunlight. But is protecting skin from the sun enough to prevent cancer? In this issue, Mitra *et al.*¹ address this question using mice that are genetically similar to humans who have red hair and fair skin or to dark-skinned or albino individuals. They find that the often deadly skin cancer melanoma occurs more frequently in 'redheaded' mice than in the other two groups, owing to mechanisms that are unrelated to exposure to ultraviolet light².

Colours of skin, eyes and hair vary widely among humans. This variation is controlled by the amount and ratio of two forms of the pigment melanin (the red–yellow pheomelanin and the brown–black eumelanin)². Both pigment types are produced by melanocytes — cells that are located in the basal layer of the skin epidermis, in hair follicles and in the uvea of the eye. Melanocyte-stimulating hormone binds to the melanocortin 1 receptor (*MC1R*)

on the surface of melanocytes, initiating a biochemical cascade that leads to increased levels of the enzymes required for eumelanin synthesis. Disturbance of the *MC1R*-mediated signalling pathway reduces total melanin production but increases the relative abundance of pheomelanins — similar to what happens when *MC1R* is not activated³.

In humans, the *MC1R* gene can show great variability (polymorphism) in sequence. Certain polymorphisms in this gene are associated with red hair colour (RHC) and so are called RHC variants. These variants result in the loss of functional *MC1R*, and individuals carrying them often have red hair, fair skin, a tendency to freckle and little ability to tan⁴.

Caucasians are generally at a higher risk of developing melanoma than are non-Caucasians, and red-haired individuals in particular are more susceptible to melanoma than are those with other hair colours^{5–7} (Fig. 1). Moreover, epidemiological analyses from pooled data sets indicate⁴ that RHC variants of *MC1R* are associated with melanoma risk. These observations are not surprising, as red hair is rich in pheomelanin and has little eumelanin. Because eumelanin shields

*This article and the paper under discussion¹ were published online on 31 October 2012.



intake of antioxidants, decrease melanoma risk in redheads. Moreover, red-haired individuals should undergo frequent dermatological skin checks, besides avoiding sun exposure. ■

Mizuho Fukunaga-Kalabis and Meenhard Herlyn are in the Tumor Microenvironment and Metastasis Program, Melanoma Research Center, The Wistar Institute, Philadelphia, Pennsylvania 19104, USA.
e-mails: mkalabis@wistar.org;
herlynm@wistar.org

1. Mitra, D. *et al.* *Nature* **491**, 449–453 (2012).
2. Hunt, G. *et al.* *Pigment Cell Res.* **8**, 202–208 (1995).
3. Sturm, R. A., Teasdale, R. D. & Box, N. F. *Gene* **277**, 49–62 (2001).
4. Raimondi, S. *et al.* *Int. J. Cancer* **122**, 2753–2760 (2008).
5. Beral, V., Evans, S., Shaw, H. & Milton, G. *Br. J. Dermatol.* **109**, 165–172 (1983).
6. Bliss, J. M. *et al.* *Int. J. Cancer* **62**, 367–376 (1995).
7. Veierød, M. B. *et al.* *J. Natl Cancer Inst.* **95**, 1530–1538 (2003).
8. Abdel-Malek, Z. A., Knittel, J., Kadekaro, A. L., Swope, V. B. & Starner, R. *Photochem. Photobiol.* **84**, 501–508 (2008).
9. Dankort, D. *et al.* *Nature Genet.* **41**, 544–552 (2009).
10. Flaherty, K. T. *et al.* *N. Engl. J. Med.* **363**, 809–819 (2010).
11. Landi, M. T. *et al.* *Science* **313**, 521–522 (2006).
12. Brenner, M. & Hearing, V. J. *Photochem. Photobiol.* **84**, 539–549 (2008).
13. Samokhvalov, A. *et al.* *Photochem. Photobiol.* **81**, 145–148 (2005).
14. Samokhvalov, A. *ChemPhysChem* **12**, 2870–2885 (2011).

Figure 1 | Genetics and the sun. Individuals with fair skin and red hair are more susceptible to melanoma than people who have darker complexions. Mitra *et al.*¹ show that this may be due not only to their greater sensitivity to sun exposure, but also to variants in their *MC1R* gene.

the redhead mice a white coat colour, it also reduced the incidence of melanoma among them. These results strongly indicate that the pheomelanin synthesis pathway plays an important part in melanoma development.

Why do redhead mice develop melanoma in the absence of UV light? It seems that redheads carrying RHC variants have a higher risk of melanoma because of intrinsic oxidative DNA damage, in addition to their poor protection from UV. Eumelanin is a strong antioxidant and reduces the accumulation of DNA damage by absorbing reactive oxygen species (ROS). Although pheomelanin may increase cancer risk by generating ROS in response to UV exposure¹², this pigment — or its chemical intermediates — can also generate ROS through a mechanism independent of UV radiation^{13,14}. The present paper shows that ROS-mediated damage to both DNA and lipids accumulates more readily in the skins of *Mc1r*-mutant redhead mice than in *Mc1r*-mutant white mice, even without UV exposure. Moreover, any such exposure seems to exacerbate oxidative damage selectively in redhead mice¹.

Studies are under way to determine whether the increased melanoma risk in redheads is limited only to cancers driven by mutant *BRAF*, or whether it also applies to other melanoma oncogenes such as *NRAS*. In addition, it remains to be confirmed whether the risk of this cancer increases in individuals whose red hair is conferred by polymorphisms in other genes of the pigment pathway.

But perhaps the most pertinent question is what can be done, beyond sun protection, to decrease melanoma risk in red-haired, fair-skinned individuals? Destroying all melanin is not an option, because the skin would lack a natural sunshield — eumelanins. Besides, Mitra and co-workers' black mice were

relatively well protected against melanoma even though they possessed both pheomelanin and eumelanin. It is conceivable that, in these animals, the abundant eumelanin scavenges pheomelanin-derived ROS. Therefore, it would be of great interest to determine whether topical compounds that induce eumelanin synthesis (such as forskolin), or oral

QUANTUM PHYSICS

Putting a spin on photon entanglement

Entanglement between a photon and a stationary particle is a key resource for quantum communication. The effect has now been observed for a photon and a single electron spin in a semiconductor nanostructure. SEE LETTERS P.421 & P.426

SOPHIA E. ECONOMOU

Quantum mechanics, the theory that explains the properties of matter through the motion of its microscopic constituents, is known for its peculiar and counter-intuitive concepts. These concepts are not only crucial to the structure of the theory, but also form the basis of many technologies, including lasers, the scanning tunnelling microscope and atomic clocks. In this issue, De Greve *et al.*¹ and Gao *et al.*² describe how they have independently demonstrated one such concept — quantum entanglement — in a semiconductor nanostructure known as a quantum dot.

Entanglement is arguably the most striking feature of quantum mechanics. Two systems, A and B, are said to be entangled if their quantum states are correlated — that is, not

separable. This means that although we cannot predict what state system A will be in when it is measured, we can predict with certainty the measurement outcome of system B once A is measured, even if the two systems are separated by an arbitrarily large distance. This feature famously unsettled Einstein — who deemed it “spooky action at a distance” — and led to the renowned ‘EPR’ paper questioning the completeness of quantum theory³.

In recent decades, with the advent of quantum information science⁴, entanglement has evolved from being considered a counter-intuitive curiosity of quantum theory to being recognized as a key resource for quantum communication and computation. In particular, entanglement between a stationary quantum system and a ‘flying’ quantum system, such as a photon, has the potential to allow long-distance,

secure communication and entanglement between two stationary systems that have never 'met' but that become linked by the flying system. Therefore, if quantum information processing is to become a practical reality, it is crucial to generate and quantify entanglement between viable physical systems. In recent years, such entanglement has been shown for trapped ions⁵ and for certain defect states in diamond called nitrogen-vacancy centres⁶.

Now, for the first time, De Greve *et al.* and Gao *et al.* demonstrate entanglement between a photon and a single electron spin trapped in a quantum dot. Formed at the interface between two semiconductors, the quantum dot has the ability to trap single electrons. Electrons confined in quantum dots remain in discrete, atomic-like spatial states without the need for external electromagnetic traps, and they have well-defined quantum properties. Once an electron is trapped in the dot, experimenters can focus on its spin, which can be in two distinct states ('up' or 'down') or, according to the superposition principle of quantum mechanics, in a linear combination of the two. In the beautiful experiments of De Greve *et al.* and Gao *et al.*, it is the electron's spin that becomes entangled with the polarization and the colour (frequency) of a photon, respectively. Polarization describes whether the light's electric field oscillates horizontally, vertically, in a circle or in any combination of these.

To generate the photon, both groups made use of the fact that quantum dots are optically active. First, the authors prepared the system using a technique called optical pumping, which amounts to depleting the spin-down state with a laser and pumping the spin into the up state. They then used a laser pulse of appropriate frequency and duration to excite the system to a higher-energy state and allowed it to 'relax' back to either one of the two lower-energy states. This kind of relaxation is accompanied by the spontaneous emission of a photon. With this process, the entire electron-photon system is in a superposition of two states: one in which the electron has spin-up orientation and the photon is blue and vertically polarized, and the other in which the electron has a spin-down orientation and the photon is red and horizontally polarized. The state of the emitted photon is completely quantum correlated (entangled) with the final electron spin state.

One problem with this kind of final state is that both the colour and polarization of the photon are correlated with the electron spin. For entanglement to be used in applications, only one of the two should be correlated. Therefore, both groups eliminated the unwanted correlation with the additional photon property. De Greve and colleagues downconverted the photon (that is, reduced its frequency by half), thereby broadening its frequency range. This step was crucial to obtaining the desired frequency state of the photon, because the correlation between electron spin and photon colour

was erased as a result of having one, broadened frequency value instead of two, distinct, sharp ones. The remaining quantum entanglement was then solely between the electron spin and the photon polarization. The downconversion has the additional benefit of providing a photon that is at the wavelength used in telecommunications, in which lossless photon propagation in optical fibres can be extremely long. As a result, this experiment is an important first step towards achieving practical quantum networks⁷ based on quantum-dot electron spins and photons.

Meanwhile, Gao and colleagues opted to use frequency as their entangled quantity, and so had to remove the correlation with polarization. They achieved this by 'filtering' the photon, regardless of its frequency, so that it acquired an anticlockwise, circular polarization. This eliminates the correlation with polarization, leaving an entangled state of photon colour and electron spin. Their filtering also solved the practical problem of separating the emitted photon from the vast number of photons emitted by the laser, which had a clockwise, circular polarization.

Using quantum dots as photon emitters has potential advantages over competing systems. Quantum dots have a higher electric dipole moment, leading to faster photon emission. What's more, they can be placed in optical microcavities to increase photon yield⁸. One avenue for future research would be the addition

of a second, distant quantum dot to the authors' set-up, with joint measurement of the two emitted photons (one from the first quantum dot and the other from the second) to entangle the spins of the two dots⁹. Another possibility would be to 're-pump' a single quantum dot to extract a second photon, which under certain conditions will be entangled with the first. This idea could be extended¹⁰ to produce multi-photon entangled states for 'measurement-based' quantum computing¹¹. The present experiments pave the way to such demonstrations, and contribute to efforts to achieve large-scale, practical processing of quantum information. ■

Sophia E. Economou is in the Naval Research Laboratory, Washington DC 20375, USA.
e-mail: sophia.economou@nrl.navy.mil

1. De Greve, K. *et al.* *Nature* **491**, 421–425 (2012).
2. Gao, W. B., Fallahi, P., Togan, E., Miguel-Sanchez, J. & Imamoglu, A. *Nature* **491**, 426–430 (2012).
3. Einstein, A., Podolsky, B. & Rosen, N. *Phys. Rev.* **47**, 777–780 (1935).
4. Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information* (Cambridge Univ. Press, 2000).
5. Blinov, B. B., Moehring, D. L., Duan, L.-M. & Monroe, C. *Nature* **428**, 153–157 (2004).
6. Togan, E. *et al.* *Nature* **466**, 730–734 (2010).
7. Cirac, J. I., Zoller, P., Kimble, H. J. & Mabuchi, H. *Phys. Rev. Lett.* **78**, 3221–3224 (1997).
8. Stute, A. *et al.* *Nature* **485**, 482–485 (2012).
9. Moehring, D. L. *et al.* *Nature* **449**, 68–71 (2007).
10. Economou, S. E., Lindner, N. & Rudolph, T. *Phys. Rev. Lett.* **105**, 093601 (2010).
11. Raussendorf, R. & Briegel, H. J. *Phys. Rev. Lett.* **86**, 5188–5191 (2001).

VASCULAR BIOLOGY

Nitric oxide caught in traffic

Nitric oxide is a vital signalling molecule that controls blood flow and pressure. Unexpectedly, a redox switch in the protein haemoglobin α within endothelial cells regulates this molecule's diffusion in blood vessels. [SEE LETTER P.473](#)

**MARK T. GLADWIN
& DANIEL B. KIM-SHAPIO**

Nitric oxide is a vasodilator. It is produced in endothelial cells, which line blood vessels, and mediates signalling cascades in adjoining smooth-muscle cells to affect the regulation of blood pressure, blood flow and oxygen delivery. Nitric oxide is proposed to control these crucial physiological processes through simple unregulated diffusion from its site of production to its target sites. But on page 473 of this issue, Straub and colleagues provide evidence¹ that the oxidation state of the protein haemoglobin α , which is expressed at the junction between endothelial and muscle cells, regulates

nitric oxide diffusion and signalling*.

Haemoglobin is best known for mediating oxygen delivery by erythrocytes (red blood cells). Nonetheless, low concentrations of this protein are expressed in other cells, such as human lung epithelial cells². In addition, over the past 15 years several primordial globins—including neuroglobin and cytoglobin—have been discovered that are expressed in various non-erythrocytic organs such as the brain, retina, endocrine organs and vascular smooth muscle. Functions that are being explored for these proteins include mediating electron-transfer reactions (such as the reduction of the enzyme cytochrome *c* and

*This article and the paper under discussion¹ were published online on 31 October 2012.

nitrite (NO_2^-) reduction) and nitric oxide (NO) scavenging reactions^{3,4}. For instance, flavohaemoglobins and myoglobin have been proposed^{4,5} to limit NO signalling by converting it to nitrate (NO_3^-) in a dioxygenation reaction. A role for haemoglobin in NO-scavenging reactions is appealing on theoretical grounds because these reactions are fast and can occur at the low globin concentrations found in cells.

The unique diffusion properties of NO (it forms a spherical concentration gradient around its source) create challenges for directional and compartmental signalling by this molecule. To reach its signalling target, NO must escape reactions that convert it to inert species.

Several solutions have evolved. These include spatial localization and control of NO production within cell-membrane structures called caveoli, and within hot spots of metabolic enzymes, where NO is coupled to its target. Alternatively, Straub *et al.* propose that the thick walls of the internal elastic lamina (the outermost elastic tissue of blood vessels that separates endothelium and smooth-muscle cells) create physical distance between the source of NO production (the enzyme endothelial NO synthase) and the NO target, the smooth-muscle soluble guanylyl cyclase enzyme.

At regular points between the endothelium and smooth muscle, known as myoendothelial junctions (MEJs), the cells project through the internal elastic lamina and 'kiss' each other, increasing the NO concentration at these corridors. These points of cell–cell contact contain gap-junction channels that allow electrical coupling and intercellular diffusion of vasodilatory factors, such as eicosanoids, potassium ions and hydrogen peroxide, and small solutes that can dynamically regulate blood-vessel diameter⁶. Straub *et al.* propose that these junctions also create corridors within the internal elastic lamina for NO diffusion.

The authors find that haemoglobin α is concentrated at the MEJs. Here, this protein serves to block NO diffusion to smooth muscle, specifically through the extremely fast and irreversible dioxygenation reaction of NO with ferrous (Fe^{2+}) oxyhaemoglobin to form nitrate and methaemoglobin, in which the iron in the haem group of haemoglobin is in the ferric (Fe^{3+}) state^{7,8}.

The team also reports that the enzyme cytochrome b5 reductase 3 — also called methaemoglobin reductase — forms a complex with cellular haemoglobin α and regulates NO diffusion by reducing the Fe^{3+} -haem to the oxygen-binding Fe^{2+} -haem. This provides an enzymatic mechanism to control NO diffusion (Fig. 1), because only the Fe^{2+} -haem can scavenge NO. By contrast, the Fe^{3+} -haem does not bind NO tightly and so allows its diffusion through the MEJ to the smooth muscle, where it activates soluble guanylyl cyclase and

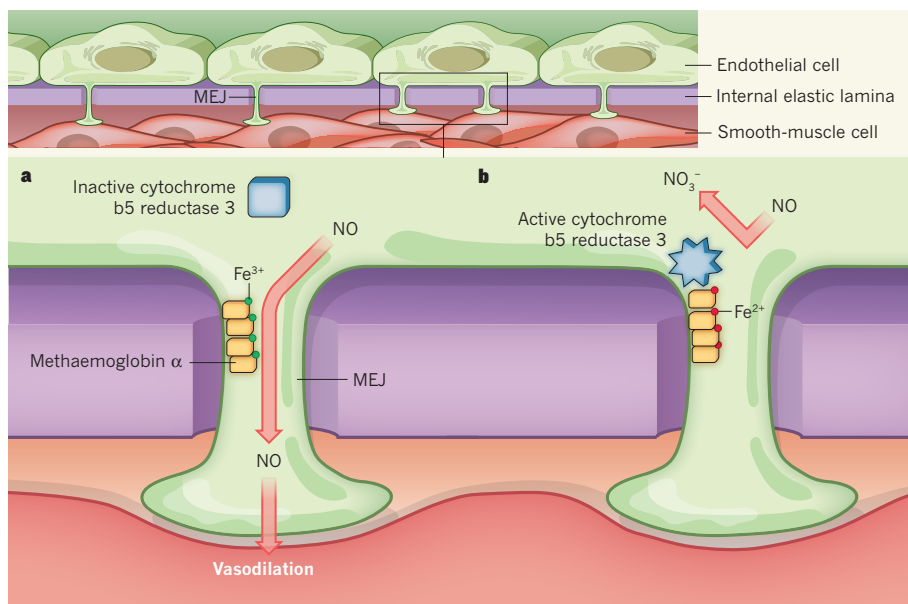


Figure 1 | Regulation of nitric oxide (NO) diffusion by cytochrome b5 reductase 3. Endothelial cells of blood vessels connect to their neighbouring muscle cells through myoendothelial junctions (MEJs), which span internal elastic lamina. **a**, Straub *et al.*¹ show that when cytochrome b5 reductase 3 is inactive, the abundant haemoglobin α at the MEJs is in the ferric (Fe^{3+}) methaemoglobin state, which does not bind tightly to the vasodilator NO. Consequently, NO passes to muscle cells where it mediates signalling cascades that result in vasodilation through activation of soluble guanylyl cyclase. **b**, Active cytochrome b5 reductase 3 forms a complex with methaemoglobin α , reducing it to the Fe^{2+} state. The latter blocks NO diffusion to smooth muscle, through a dioxygenation reaction that results in the formation of nitrate (NO_3^-).

mediates the downstream signalling cascade.

Another outcome of NO interaction with methaemoglobin is reductive nitrosylation, which can form the signalling molecules nitrite or S-nitrosothiols⁹. Both nitrite and S-nitrosothiols can regulate NO signalling independently of soluble guanylyl cyclase and by post-translational modification of target proteins. However, the reductive nitrosylation reaction is some 200 times slower than the scavenging reactions⁹. Nonetheless, the Fe^{3+} -methaemoglobin allows NO to diffuse through the MEJ or to react and form the freely diffusible nitrite and S-nitrosothiols.

Straub and colleagues' work complements other studies^{3,10,11} that were conducted under conditions of hypoxia (oxygen shortage) and which found that deoxygenated globins can function as nitrite reductase enzymes that react with nitrite to generate NO. For instance, when oxygen levels are low in smooth-muscle cells, myoglobin can reduce nitrite to NO, contributing to vasodilation¹¹. In this reaction, the Fe^{2+} -deoxymyoglobin transfers an electron to nitrite to form NO and metmyoglobin. It seems, therefore, that globins can limit NO signalling when reduced and oxygenated and enhance NO signalling when oxidized and deoxygenated (Fig. 1).

This paper highlights a novel function of the MEJ as an NO diffusion corridor. The expression and localization of haemoglobin α chains and cytochrome b5 reductase 3 at the MEJ constitute a specific checkpoint or traffic light for redox-regulated NO diffusion at these corridors. Future work must clarify

the post-translational modifications of cytochrome b5 reductase 3 that control its activity, and which would ideally couple activation of NO synthase to NO diffusion and MEJ 'gate opening'. Moreover, haemoglobin functions beyond NO scavenging must be explored, not least because of the widespread expression of cellular haemoglobins in plants, which do not possess NO synthase enzymes. ■

Mark T. Gladwin is at the Vascular Medicine Institute and the Division of Pulmonary, Allergy and Critical Care Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania 15213, USA. **Daniel B. Kim-Shapiro** is in the Department of Physics and the Translational Science Center, Wake Forest University, Winston-Salem, North Carolina 27109, USA. e-mail: gladwinmt@upmc.edu

1. Straub, A. C. *et al.* *Nature* **491**, 473–477 (2012).
2. Newton, D. A., Rao, K. M., Dluhy, R. A. & Baatz, J. E. *J. Biol. Chem.* **281**, 5668–5676 (2006).
3. Tiso, M. *et al.* *J. Biol. Chem.* **286**, 18277–18289 (2011).
4. Ascenzi, P. & Brunori, M. *Biochem. Mol. Biol. Educ.* **29**, 183–185 (2001).
5. Gardner, P. R., Gardner, A. M., Martin, L. A. & Salzman, A. L. *Proc. Natl Acad. Sci. USA* **95**, 10378–10383 (1998).
6. Segal, S. S. & Bagher, P. *Circ. Res.* **106**, 1014–1016 (2010).
7. Eich, R. F. *et al.* *Biochemistry* **35**, 6976–6983 (1996).
8. Gladwin, M. T. *et al.* *Proc. Natl Acad. Sci. USA* **97**, 9943–9948 (2000).
9. Tejero, J. *et al.* *J. Biol. Chem.* **287**, 18262–18274 (2012).
10. Cosby, K. *et al.* *Nature Med.* **9**, 1498–1505 (2003).
11. Totzeck, M. *et al.* *Circulation* **126**, 325–334 (2012).



Cover illustration
Nik Spencer

Editor, *Nature*
Philip Campbell

Publishing
Nick Campbell
Claudia Deasy

Production Editor
Jenny Rooke

Art Editor
Nik Spencer

Sponsorship
Reya Silao

Production
Emilia Orviss

Marketing
Elena Woodstock
Hannah Phipps

Editorial Assistant
Anastasia Panoutsou

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel: +44 (0) 20 7833 4000
e: nature@nature.com



nature publishing group

Abnormal metabolism is at the heart of some serious health problems (such as obesity, diabetes and cancer), which not only reduce our life expectancy, but are also a great cost to society. This Insight offers a snapshot of the molecular mechanisms that underlie metabolism and its associated pathology, and showcases the progress made in this buoyant area of research.

Metabolism beats to a drum of about 24 hours. In his Review, Joseph Bass shows how breakthroughs in our understanding of circadian rhythms and molecular clocks bring insight to the molecular pathogenesis of metabolic disorders.

Although direct actions of hormones, such as insulin, regulate nutrient handling in peripheral tissues, the central nervous system also plays a significant part in metabolic regulation. From specific signalling mechanisms to hypothalamic circuitry, Martin Myers and David Olson discuss how the brain controls metabolism.

Cancer cells thrive by switching to a different metabolic program. The mechanistic basis for these changes and how they are connected to oncogenic pathways is becoming increasingly understood. Almut Schulze and Adrian Harris discuss these advances, and explore strategies that interfere with these metabolic pathways for use as anticancer therapies.

Much of our understanding of mitochondria has come from studying rare mitochondrial disorders. Scott Vafai and Vamsi Mootha review what we have learned from these diseases by putting them in the context of a contemporary understanding of mitochondrial evolution, biochemistry and genetics.

Finally, Jeremy Nicholson and colleagues discuss how metabolic phenotyping, which involves a comprehensive analysis of biological fluids or tissue samples, can facilitate the biochemical classification of individual physiological or pathological states. This approach is currently being used in clinical practice to assist with disease diagnosis, prognosis and treatment selection for individual patients and to estimate disease risks at the population level.

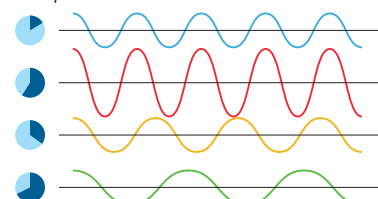
**Joshua Finkelstein, Noah Gray,
Marie Thérèse Heemels, Barbara Marte & Deepa Nath**
Senior Editors

CONTENTS

REVIEWS

348 Circadian topology of metabolism

Joseph Bass



357 Central nervous system control of metabolism

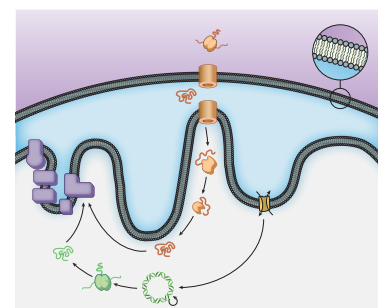
Martin G. Myers Jr & David P. Olson

364 How cancer metabolism is tuned for proliferation and vulnerable to disruption

Almut Schulze & Adrian L. Harris

374 Mitochondrial disorders as windows into an ancient organelle

Scott B. Vafai & Vamsi K. Mootha

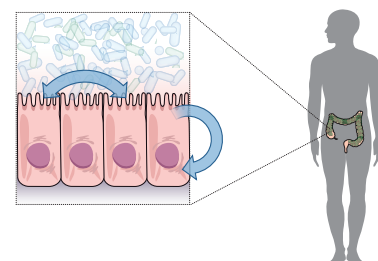


384 Metabolic phenotyping in clinical and surgical environments

Jeremy K. Nicholson, Elaine

Holmes, James M. Kinross, Ara W.

Darzi, Zoltan Takats & John C. Lindon



Circadian topology of metabolism

Joseph Bass^{1,2}

Biological clocks are genetically encoded oscillators that allow organisms to anticipate changes in the light–dark environment that are tied to the rotation of Earth. Clocks enhance fitness and growth in prokaryotes, and they are expressed throughout the central nervous system and peripheral tissues of multicelled organisms in which they influence sleep, arousal, feeding and metabolism. Biological clocks capture the imagination because of their tie to geophysical time, and tools are now in hand to analyse their function in health and disease at the cellular and molecular level.

Benjamin Franklin's dictum "early to bed, early to rise" is built on the supposition that sleep is inevitable at night and waking should correspond to sunrise. However, some people are 'larks' and wake early, whereas others are 'night owls' and stay up late, hinting that there is a biological driver of sleep–wake rhythms. A triumph of modern genetics has been the identification of the molecular pathways that dictate the sleep–wake cycle and other 24-hour-circadian (derived from *circa diem*, about a day) rhythms (Fig. 1). Many of the principles of this system originally came from genetic studies of model-organism mutants with altered period phenotypes, and are described in this Review in the context of our current understanding of clock systems in mammalian tissues. These advances in understanding can also be considered in the context of human studies that suggest that artificial light, night work, a reduction in normal sleep time, shift work, travel and temporal disorganization, all of which are common in industrialized societies, have disrupted the pattern of alignment between the external light–dark cycle and the internal clock, which was set to a 24-hour day early in evolution (Fig. 2). Identification of the molecular clock may lead to insight into circadian and sleep disorders in humans. This Review highlights how advances in the field of molecular clocks could help in understanding the molecular pathogenesis of metabolic disorders across the lifespan.

Origins of circadian clocks

Consciousness of the temporal world has been a hallmark of civilization from prehistoric times — reflected in the iconic solar worship site at Stonehenge. Yet, the correspondence between biological and geophysical phenomena was not recognized until the early 1700s, when the French astronomer Jean-Jacques d'Ortous De Mairan demonstrated that the leaves of *Mimosa pudica* continue to open and close every 24 hours even when the plant was enclosed in a sealed box. Since then, other photosynthetic organisms, and nearly all forms of life on the surface of the planet, have been shown to exhibit similar circadian cycles. The principal criteria of circadian oscillators emerged from the work of Pittendrigh and Aschoff, who established the defining characteristics of biological clocks: a persistent and sustained period length under constant conditions, entrainment to environmental signals such as light, and stability across wide variations in temperature (referred to as temperature compensation).

The transcriptional motif

A leap forward in our understanding of molecular clocks came from the deliberate mutagenesis studies of Konopka and Benzer in the fruitfly *Drosophila melanogaster*. These studies screened for 24-hour rhythms

in the fly's emergence from the pupal case, and provided a gateway to the modern era of circadian genetics¹. A key advance in circadian genetics was the concept that clocks comprise a transcription autoregulatory feedback loop, with the forward limb encoding activators that promote transcription of a set of repressors, which feed back to inhibit expression and function — a cycle that repeats itself every 24 hours across divergent phyla^{2,3} (Fig. 1). Indeed, transcriptional oscillators may have provided a selective advantage early in evolution by averting the DNA-damaging effects of sunlight. The presence of a photolyase domain in clock repressors indicates that the timing systems co-evolved with DNA repair⁴.

A metabolic variation on transcription

Although the transcription feedback loop represents a conserved model of the circadian oscillator, recent advances point towards metabolic oscillators as an additional mechanism of circadian timekeeping that can be, in certain circumstances, independent of transcription. A remarkable series of test-tube experiments^{5,6} has provided the most convincing evidence for a protein-based clock. A complex of just three proteins (KaiA, KaiB and KaiC) together with ATP undergo a self-sustaining 24-hour cycle of alternating phosphorylation and dephosphorylation. The phosphorylation cycle is both coupled to and directs gene transcription, although it can also be seen in the absence of transcription under certain circumstances⁵. The cycling of this kinase–phosphatase reaction remains constant at different temperatures — a defining feature of circadian oscillators⁶ (Fig. 3). The turnover of ATP modulates the Kai phosphorylation cycle, suggesting that the clock may be coupled to metabolic activity⁷. An interdependence between circadian and metabolic oscillators has also been suggested by showing that the activity of clock transcription factors is sensitive to redox state⁸ (Fig. 3).

More recently, periodic flux in metabolic cycles has been related to production of reactive oxygen species (ROS). Peroxiredoxin is a redox-sensitive protein, which has a reactive thiol within the active site that is involved in electron transfer from reactive oxygen. These proteins exhibit 24-hour oscillation in cells in the absence of transcription — as shown in erythrocytes⁹ — and under conditions when transcription is arrested in the alga *Ostreococcus tauri*¹⁰. Peroxiredoxin proteins have been shown to exhibit self-sustained oscillation in many species of archaeobacteria, plants and in other eukaryotic cells, suggesting that a redox cycle may be one of the most conserved 24-hour oscillators¹¹ (Fig. 3). Determining whether the peroxiredoxin proteins are themselves clocks or reporters of a more primary metabolic oscillator will require further genetic and biochemical analyses.

Given that eukaryotes first diverged from bacteria around 1.5 billion years ago, a provocative speculation is that oxygenation

¹Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois 60611, USA. ²Department of Neurobiology, Northwestern University, Weinberg College of Arts and Sciences, Evanston, Illinois 60208, USA.

of the atmosphere conferred an adaptive advantage on organisms with a redox-based clock. If this speculation is true, then organisms existing in anaerobic conditions, such as deep-sea vaults, might have less pressure to eliminate ROS, and may be devoid of a metabolic clock. In mammals, oscillations in the peroxiredoxin redox state have been proposed to represent a means of rhythmically anticipating the generation of ROS. Conceivably, organisms that are capable of efficiently extinguishing ROS may have had a survival advantage during the oxygen expansion of the atmosphere. If this is true, then the peroxiredoxin clock may keep time at the organismal level by regulating the oscillation of ROS. Consideration of the origins of internal clocks remains central to our understanding of links between circadian and metabolic systems (Fig. 3): was it escape from sunlight, avoidance of toxic metabolites during respiration or elimination of inefficient metabolic cycles that drove these processes together? Identifying metabolic circadian clocks could determine whether there is a monophyletic origin of organisms that possess such oscillators, emphasizing the fundamental clues that clocks may yield for understanding evolutionary and functional relationships among species.

Clocks impact on fitness in multicellular organisms

A principle to emerge from genetic studies is that period length is genetically programmed. But how and why are the body's internal clocks set to 24 hours? Resonance studies are perhaps the best approach to address this question because they discriminate between oscillator-dependent and independent gene functions. Phenotypes such as growth and reproduction are monitored under conditions in which alignment between internal period length and the environmental light cycle are systematically varied (long- and short-period mutants are maintained under lengthened or shortened light cycles). In bacteria and plants, such studies suggest an advantage to period alignment¹². But, what are the consequences of misalignment at the molecular level? One possibility is that misalignment reduces genome stability by shuffling the phase relationship between cycles of DNA damage and repair. Alternatively, misalignment may superimpose incompatible biochemical processes, such as the oxidative and reductive phases of the metabolic cycle. For instance, in yeast (*Saccharomyces cerevisiae*), the mutation rate increases when metabolic cycles are misaligned with DNA replication¹³. Furthermore, although circadian cycles are directly photosensitive in plants and even in the cells of *D. melanogaster*, in mammals the role of cryptochromes as both a timekeeper and a DNA-damage-repair agent is uncoupled. It is tempting to speculate that the coupling of DNA repair to circadian cycles may contribute to ageing even in higher eukaryotes.

Neural-clock sensory circuit and circadian-system ageing

Understanding how complex organisms detect light and synchronize the clocks in brain and other tissues to the environment remains a central challenge in circadian research. Sensory pathways within the brain have been identified that synchronize the clock independently of visual image formation. The light-response pathway also synchronizes neural and peripheral clocks, but the integrity of the clock network may decline with age.

Light-response pathway in the retina

A breakthrough in understanding the entrainment of oscillators in animals came with the discovery that even mice without the classic rod and cone visual photoreceptors are still able to synchronize their internal clocks to light. The light photoreceptors are expressed in a small number of retinal cells that express the photopigment melanopsin¹⁴. Mice that are genetically depleted of these few hundred cells still have normal vision, but they are unable to synchronize their clocks to light¹⁵. All connections to the hypothalamic pacemaker neurons in the suprachiasmatic nucleus go through these few melanopsin cells¹⁶. Suprachiasmatic-nucleus neurons comprise

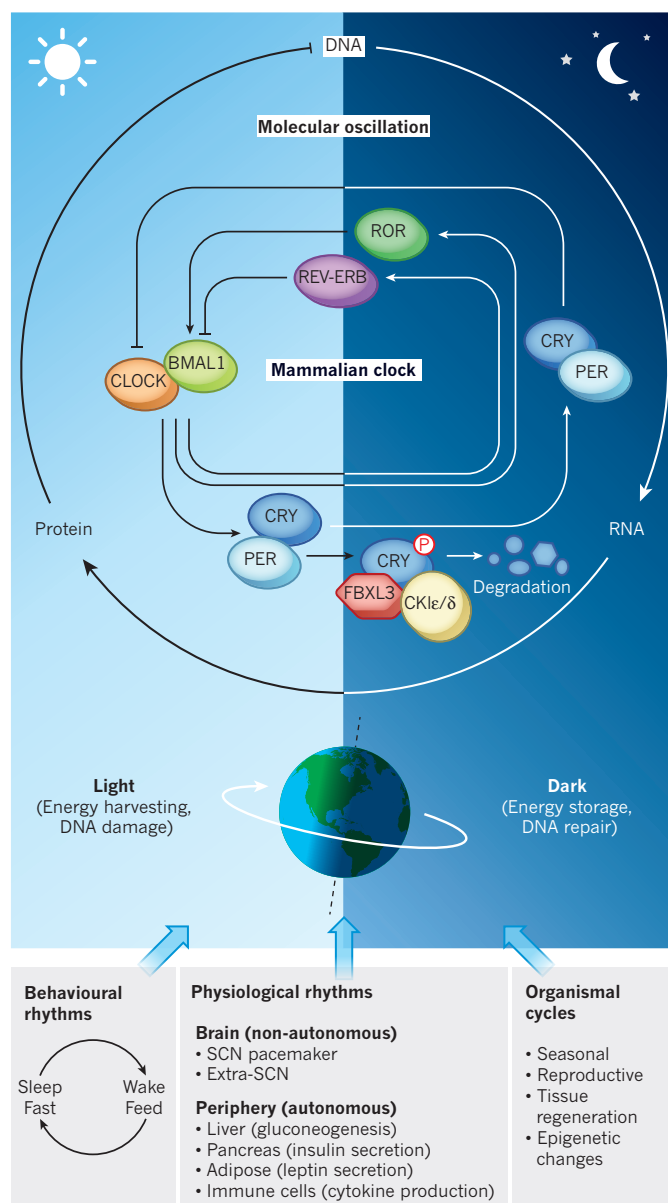


Figure 1 | Circadian adaptation as a unifying model that integrates behaviour and physiology. The circadian clock allows light-sensitive organisms to synchronize their daily molecular oscillations, behavioural rhythms, physiological rhythms and organismal cycles with the rotation of Earth on its axis. Core molecular pathways dictate behavioural and physiological cycles. This core molecular clock in mammals, expressed both in brain and peripheral metabolic tissues, comprises a series of transcription-translation feedback loops that include opposing transcriptional activators (CLOCK-BMAL1) and repressors (PER-CRY)¹. The non-phosphorylated PER-CRY complex represses CLOCK-BMAL1; phosphorylation, in turn, results in the degradation of PER-CRY and the turnover of these repressors. In addition, CLOCK-BMAL1 induces transcription of REV-ERB and of ROR, which regulate BMAL1 expression. During the night, PER-CRY is degraded through the ubiquitination of CRY by FBXL3. The circadian clock coordinates anabolic and catabolic processes in peripheral tissues with the daily behavioural cycles of sleep-wake and fasting-feeding. SCN, suprachiasmatic nucleus.

the central node in the clock network, which in turn synchronizes the hypothalamic control of energy balance, sympathetic outflow and the neuroendocrine systems¹⁷. A future goal will be to extend studies of the circadian neurosensory circuit to understand whether factors that affect sensory perception or nutrition in early development might alter synaptology within this circuit.

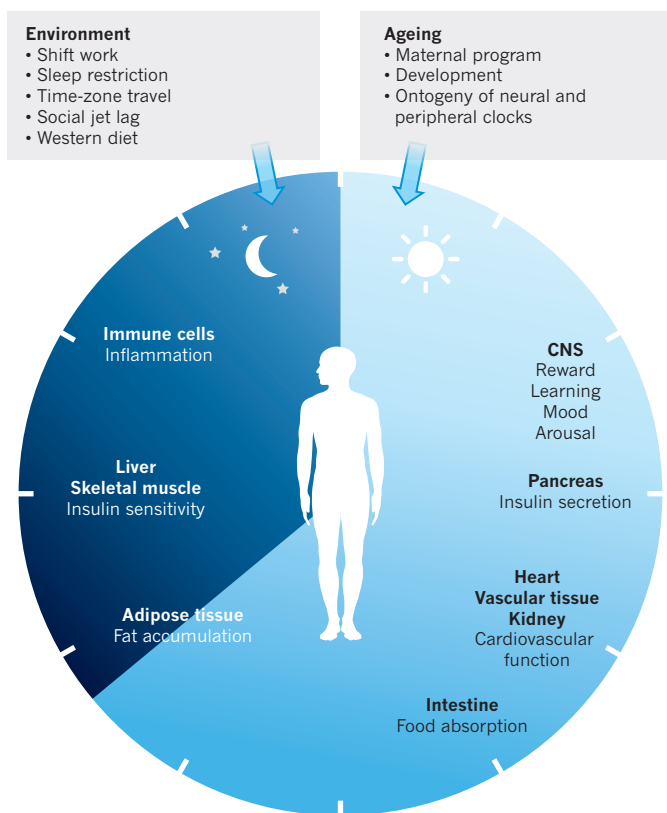


Figure 2 | Affect of ageing and environmental disruption on circadian control of metabolic processes. The circadian clock partitions metabolic processes within the peripheral tissues according to whether we are asleep or awake; for example, the pancreatic clock promotes insulin secretion during the wake–feeding period⁵², but the adipose tissue clock promotes fat accumulation during the sleep as well as the wake period. Synchronization of peripheral tissue clocks and downstream metabolic processes with the environmental cycle is crucial for the maintenance of the health of the organism^{35,39}. We are only just beginning to gain an appreciation of how both ageing^{26,28} and environmental disruption (including changes in diet, time of feeding or jet lag) perturb the integration of the circadian and metabolic networks¹⁰⁰. CNS, central nervous system.

Hypothalamic networks link circadian and energetic centres

The next question is how does the hypothalamic clock communicate with extra-pacemaker and peripheral tissues to produce a coherent phase in the circadian systems throughout the organism? An appreciation of the central role of the suprachiasmatic nucleus in coordinating sleep–wake behaviour, as well as physiological systems, pre-dated molecular advances in defining the molecular mechanisms of the clock. Master pacemaker cells were first localized to the anterior hypothalamus¹⁸. Transplantation of the suprachiasmatic nucleus with tissue from short-period mutants into animals with damage to the anterior hypothalamus imposed short periods on the host, proving that the suprachiasmatic nucleus is a master clock, not just a relay station¹⁹. These studies also suggested that secreted factors from the suprachiasmatic nucleus contribute to the synchronization of clocks, although the identity of these synchronizing factors has been elusive.

Among the areas receiving suprachiasmatic-nucleus projections, a large output is toward the dorsal medial hypothalamic (DMH) area²⁰. This circuit has been implicated in a phenomenon known as food anticipatory activity (FAA), whereby animals increase their activity in response to food provided during the incorrect circadian phase, although it is possible that other hypothalamic regions may be necessary for FAA²¹. Moreover, the role of clock genes in FAA is controversial. Animals deficient in MC3R, a metabolic signalling receptor, have diminished FAA²², suggesting that metabolic rather than clock factors may be the principal driver of FAA.

In addition to output to the DMH area, suprachiasmatic-nucleus projections synapse on orexin-expressing neurons within the lateral hypothalamic area. Orexin (also known as hypocretin) is a neuropeptide that has been found to stimulate arousal and increase energy expenditure with intracerebroventricular administration. Deficiency of either orexin or its receptor is a hallmark of both canine genetic²³ and human autoimmune forms of narcolepsy. Interestingly, narcolepsy is also correlated with elevated body-mass index²⁴. Furthermore, ablation of orexin receptors increased susceptibility to diet-induced obesity, suggesting that the physiological role of the orexins is to promote arousal and antagonize weight gain²⁵.

Ageing and the circadian system

Given the extensive integration of the neuroendocrine and circadian systems, it is intriguing to note that suprachiasmatic-nucleus function declines with age²⁶ (Fig. 2). What molecular mechanisms might account for the observed effects of ageing on the integrity of circadian systems? Studies in circadian-clock-mutant animals have shown the susceptibility of certain tissues to damage, such as accelerated cataract formation and dermatitis²⁷. In addition, *Bmal1*-knockout mice have a premature death, compared with control mice, that is correlated with increased accumulation of ROS²⁸. Deficiency of cryptochrome, a repressor of the internal clock repressor, has also been associated with alterations in liver regeneration, emphasizing the coupling of circadian and cell-cycle pathways²⁹. Although epidemiological evidence suggests there is a link between circadian disruption and cancer risk, a full understanding of the role of circadian systems in tumorigenesis remains an area for investigation. Given the emerging link between metabolic flux and cancer-cell survival, it will be especially interesting to determine whether circadian systems also intersect with oncogenic pathways through regulation of fuel selection (discussed later).

Circadian origins of metabolic disease

An emerging theme in both circadian and metabolic studies is that it is not only the central nervous system, but also the peripheral tissues that modulate sensory response to the environment. A key goal is to determine how the circadian cycle modulates homeostatic and nutrient responsive pathways to coordinate behaviour and energetics.

Feeding resets peripheral clocks

Shifting food availability to the incorrect circadian time (through restricted feeding) in animals entrained to a standard light–dark schedule causes a shift in the peripheral clock in the liver without altering pacemaker neuron clocks or overall behavioural rhythm³⁰. One pathway that entrains the liver clock to feeding involves glucocorticoid signalling³¹. Temperature also entrains the clock through mechanisms involving transcriptional control of heat-shock factor protein 1 (refs 32 and 33) (Fig. 3). ADP-ribosylation also affects entrainment of the liver to feeding³⁴. Together, these findings suggest that multiple levels of transcriptional regulation reinforce signalling from the neuronal pacemaker to the liver, with the net result of suprachiasmatic-nucleus input restraining the response of circulating hormonal or metabolite signals that are related to feeding.

Meal timing affects metabolic syndrome

Mounting evidence suggests that alignment between central behavioural rhythms and feeding time is important in metabolic health. Animals fed a high-fat diet shift their pattern of food intake and consume nearly all of the excess calories at the incorrect circadian time (during the rest period)³⁵. A similar erosion of the partitioning of feeding between activity and rest is observed in *Clock*-mutant³⁶ and *Npas2*-knockout³⁷ mice, and a high-fat diet induces period lengthening in wild-type mice, which is a core property of the clock³⁵. Similarly, mice fed a high-fat diet exclusively during the rest period have accelerated weight gain compared with animals fed during the correct circadian time³⁸, whereas restricting access to high-fat food ameliorates diet-induced metabolic syndrome in

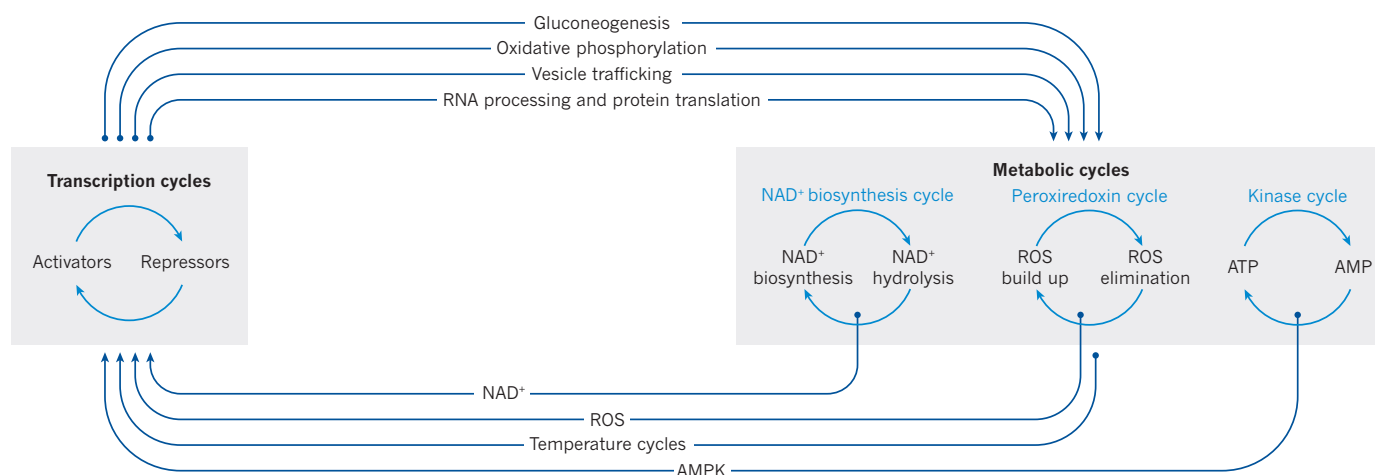


Figure 3 | Cross-talk between clock transcription and metabolic systems at the molecular and physiological levels. Cycling of circadian transcriptional activators and repressors^{2,3} controls fundamental physiological cellular and metabolic processes, including gluconeogenesis, oxidative phosphorylation, RNA processing and translation, and vesicle trafficking⁶⁴. This can occur

in various tissue types, but there may be tissue-specific differences. In turn, metabolic cycles reciprocally affect the clock: the NAD⁺ biosynthesis^{81,82}, peroxiredoxin¹¹ and kinase⁶ cycles generate active intermediates that, along with cycles of temperature, provide feedback to regulate the core clock transcriptional network. AMPK, AMP-activated protein kinase; ROS, reactive oxygen species.

mice³⁹. Even constant exposure to light causes increased insulin resistance⁴⁰. An important goal will be to elucidate the neural and molecular basis of the links between altered timing and behavioural and metabolic disruption. Conceivably, alterations in the time of feeding may induce desynchrony between suprachiasmatic-nucleus firing rhythms and input from peripheral feeding responsive signals. Altered excitability and activity of certain cell groups may occur within restricted windows as a result of either cell-autonomous clock function or non-autonomous input from the suprachiasmatic nucleus. Although much of the experimental evidence that links timing to metabolism has emerged in rodent studies, both epidemiological and clinical investigations suggest parallel mechanisms may predispose humans to metabolic pathologies^{41,42}.

Function of clock genes in metabolic and vascular disease

Genetic tools to perturb the internal clock have created opportunities to analyse the molecular basis of the clustering of certain pathologies within limited time windows, including morning myocardial infarction and hypertensive crises⁴³ (Fig. 2). The basis of increased risk of myocardial infarction in the morning is probably multifactorial; circadian transcription factors have been shown to trans activate the promoter of the pro-thrombotic cytokine plasminogen activator inhibitor type 1 (PAI-1) through both REV-ERB and E-box motifs, corresponding to oscillation in thrombosis risk^{44,45}. Arrhythmogenesis also occurs more frequently in the morning, and is associated with the control of potassium-channel expression by myocardial clock genes⁴⁶. Cardiac clocks also influence myocardial contractility and oxidative metabolism⁴⁷. Both autonomous and non-autonomous vascular effects of the clock cause variation in blood pressure across the light–dark cycle⁴⁸. CLOCK expression in the vasculature affects the progression of atherosclerosis in xenograft models⁴⁹. Similarly, dyslipidaemia arises in circadian mutants³⁶ and may be related to dysregulation of core clock genes⁵⁰ and the clock-controlled gene *Ccrn4l* (ref. 51).

Phase and tissue clocks determine physiological outcomes

Although the aforementioned studies have analysed mutations in multiple tissues, one complexity of the circadian system is that the set of clock transcriptional activators and repressors exerts opposing effects within different tissues at different times in the 24-hour cycle. This is most clearly evident in studies of glucose metabolism. For example, *Clock*-mutant mice are hyperglycaemic and have increased susceptibility to diet-induced obesity early in life, but also become hypoinsulinaemic with age³⁶. An explanation for this finding was provided by studies in which animals with selective ablation of the

circadian gene *Bmal1* within the endocrine pancreas were shown to have much more profound hyperglycaemia and β -cell failure than the multitissue mutants⁵². Indeed, whole body loss of *Bmal1* increases insulin sensitivity⁵³, and selective *Bmal1* ablation in the liver causes hypoglycaemia⁵⁴. Taken together, these findings suggest that CLOCK–BMAL1 exerts opposing effects in the liver and the pancreas — within the liver, it promotes fuel mobilization during fasting, but in the pancreas this complex promotes post-prandial insulin exocytosis. The loss of the CLOCK–BMAL1 repressor CRY results in the opposite phenotype in the liver, with increased gluconeogenesis and enhanced responsiveness to both glucagon and glucocorticoids^{55,56}.

One challenge to studies of genetically modified circadian-mutant mice is whether the observed abnormalities are dependent on alterations in the core timing process or in downstream processes independent of the internal clock. Furthermore, because ablation of individual clock factors causes compensatory changes in the expression of other factors within the core loop — even when studying effects of individual gene manipulations — both loss and gain of function analyses become important to determine whether the effects are direct or indirect⁵⁷. Studies in both activator and repressor mutants, and analyses across different circadian phases in synchronized tissues, will ultimately elucidate the link between gene function, timing and physiology.

Coupling of circadian and metabolic cycles

Extensive cross-talk between metabolic and circadian systems allows organisms not only to anticipate physiological needs in advance of the daily light–dark cycle, but also to adjust the phase of internal cycles in response to changes in the environment. The cross-talk between cellular oscillators and metabolic systems can be traced to overlapping transcriptional networks that encode circadian and nuclear-receptor signalling pathways (Fig. 4).

Transcription factors respond to metabolic flux

At the molecular level, multiple sequence alignment of core clock genes originally demonstrated the presence of the Per–Arnt–Sim (PAS) domain in the internal clock transcription factors. This conserved motif is also found within xenobiotic and hypoxaemic response transcription factors in addition to kinases and ion channels. PAS domains participate in both protein–protein interactions and direct detection of small molecules. Haem, a gas-responsive cofactor, binds to the PAS domain in a subset of these proteins and detects both carbon monoxide in the clock factor NPAS2 and oxygen in the bacterial PAS-containing histidine kinase^{58,59} (Fig. 4). Haem has also been found to bind to the

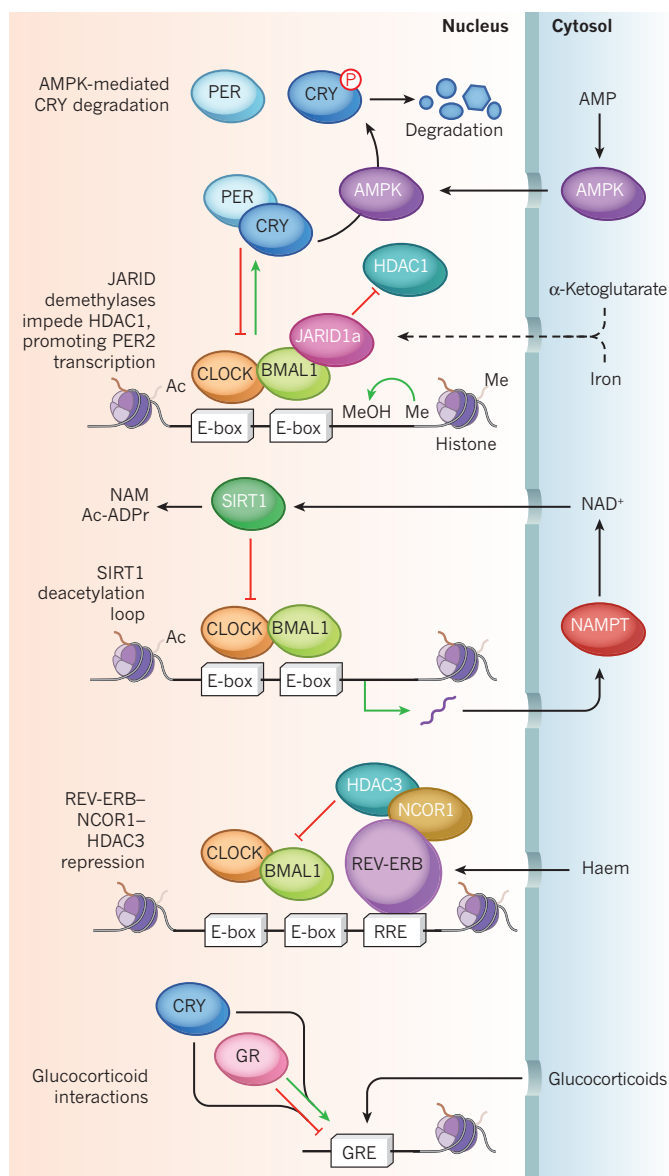


Figure 4 | Genomic and epigenetic links between circadian and metabolic systems. Research has highlighted the existence of a highly dynamic and multi-layered network of factors involved in epigenetic transitions across the circadian cycle. For example, AMP-activated protein kinase (AMPK)-mediated phosphorylation of CRY⁶³ controls proteolytic degradation of the negative arm of the central oscillator. Distinct from the 'core' oscillation shown in Fig. 1, which uses CKI to phosphorylate and therefore cause the degradation of CRY, AMPK entrains the circadian clock to the metabolic environment by using the same modification as CKI. The histone demethylase JARID1a (ref. 89) is involved in the activity of CLOCK–BMAL1 complexes (which regulate PER and CRY transcription) that are bound to E-box motifs. This activates the positive arm of the central oscillator by blocking HDAC1 activity, a function that may be tied to the use of α -ketoglutarate and iron, to hydroxylate and remove H3K4 methyl marks on histones. However, this demethylation activity seems to be dispensable in circadian regulation. NAD⁺-dependent enzyme SIRT1 activity is dependent on NAD⁺-regulating nicotinamide mononucleotide phosphoribosyltransferase (NAMPT) (a transcriptional product of the CLOCK–BMAL1 complex). SIRT1 in turn feeds back to inhibit the CLOCK–BMAL1 complex^{79,80}. Deacetylation by SIRT1 generates nicotinamide (NAM) and O-acetyl-ADP-ribose (Ac-ADPr) as by-products. The REV-ERB–HDAC3–NCOR1 repressive complex is sensitive to haem levels, possibly activating the repressor activity^{61,62}, and binds directly to the promoters of clock-controlled genes (RRE), while binding directly to CLOCK–BMAL1 deacetylating nearby histones (not shown). Transcriptional activation and repression of glucocorticoid-responsive elements (GREs) by the glucocorticoid receptor (GR) are modulated by glucocorticoids and CRY⁵⁶.

PAS domain of the nuclear hormone receptor co-repressor proteins REV-ERB- α and REV-ERB- β (refs 60–62).

Although the hallmark of the circadian cycle is its self-sustained activity, changes in circadian gene transcription in response to energetic flux have also been demonstrated. One example of a chemical signal that couples internal clock function to nutrient state involves AMP kinase (AMPK), a protein that is activated following intracellular nutrient restriction. Stimulation of AMPK leads to the phosphorylation and subsequent proteasomal degradation of the repressor CRY⁶³. An important question to ask is whether alterations in AMPK signalling modulate the output of pacemaker neurons to control behavioural and physiological rhythms, or whether AMPK signalling affects circadian transcriptional cycles in a restricted number of peripheral tissues.

Nuclear receptors connect endocrine and circadian physiology

Core clock genes can be defined as those that are necessary for circadian behavioural rhythmicity; however, it is intriguing to note that data mining suggests that a large fraction of the transcriptome exhibits circadian oscillation in tissues such as the liver, although the specific number of transcripts designated as oscillating depends on the threshold set for gene amplitude during bioinformatic extraction⁶⁴. Indeed, many of the nuclear hormone receptors in mice exhibit circadian oscillation and, in turn, the timing of interaction between nuclear receptors and ligands may be considered a broad feature of the coupling between temporal and physiological systems⁶⁵. The coupling of circadian and nuclear hormone receptor networks is reflected in the overlap between genome occupancy of the orphan nuclear receptor ERR- α and of BMAL1 (ref. 66), and in altered circadian function in mice that are deficient in the co-activator PGC-1 α (ref. 67). Surprisingly, the circadian repressor PER2 binds to several nuclear hormone receptors⁶⁸, whereas CRY binds to the glucocorticoid receptor and modulates its function⁵⁶. Glucocorticoid has also been shown to modulate the expression of REV-ERB- α (ref. 69), a regulator of *BMAL1* (also known as *ARNTL*) transcription⁷⁰. REV-ERB- α agonists have been shown to influence circadian function and may protect against adverse metabolic consequences of diet-induced obesity⁷¹. Genetic studies in which REV-ERB is eliminated only in adults have reinforced the link between REV-ERB- α and its homologue REV-ERB- β in the co-regulation of circadian and metabolic homeostasis^{72,73}.

Genomic analyses have also begun to increase the evidence for co-repressors in coupling circadian and metabolic systems (Fig. 4). HDAC3, a class I histone deacetylase that is recruited to the REV-ERB- α –NCOR1 transcriptional complex⁷⁴, exhibits diurnal occupancy of the loci that encode the genes involved in lipogenesis and carbohydrate metabolism⁷⁵. REV-ERB- α and REV-ERB- β occupy a spectrum of lipogenic genes in addition to core clock genes⁷³, although it remains unclear whether REV-ERB exerts the same effect on both classes of transcripts. For instance, BMAL1 occupies loci, encoding both internal clock and metabolic genes throughout the genome⁷⁶. However, analysis of expressed transcripts indicates closer phase alignment between genome occupancy and RNA transcription for core clock genes. Moreover, although clock genes occupy many sites throughout the genome, the core clock genes generally exhibit greater variation in expression across the cycle compared with output genes. One exception is the very high amplitude rhythm of *DBP* and *TEF*, two output genes belonging to the proline and acidic amino-acid-rich basic leucine zipper (PAR-bZIP) family of transcription factors.

Although genomic analyses bring greater focus to studies of transcriptional cross-talk, future transcriptome analyses will be necessary to determine whether occupancy of loci throughout the genome is accompanied by changes in expression. Eventually, it will be important to extend chromatin immunoprecipitation and sequencing analyses to comparison of the transcriptome in wild-type and knockout mice for the factors in question to assess function. For instance, although steatosis is observed with hepatic ablation of both HDAC3 (ref. 74) and REV-ERB- α

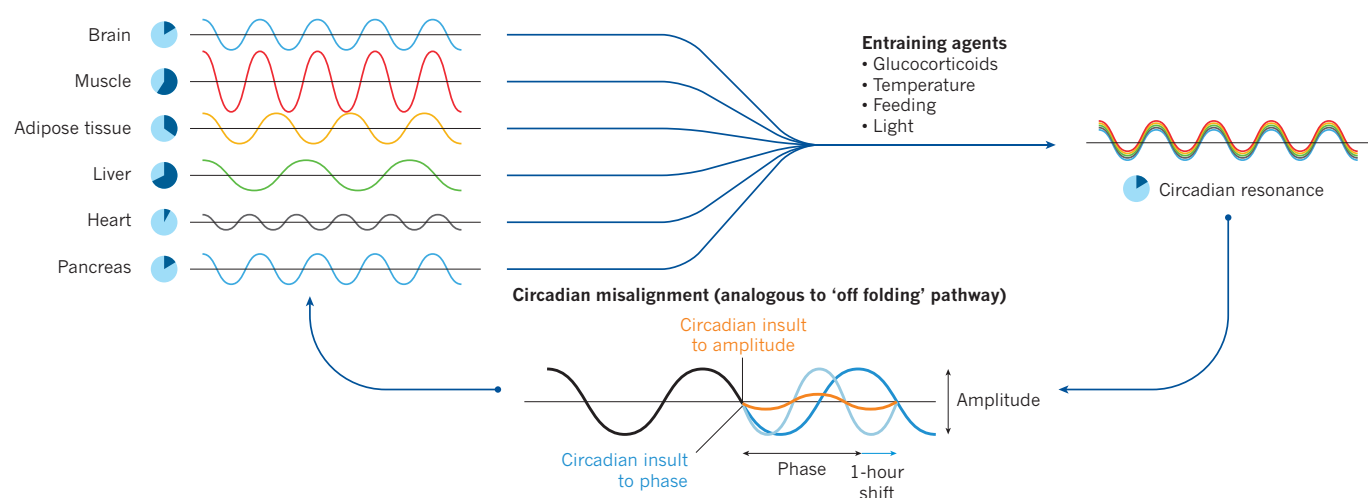


Figure 5 | Topological model of circadian physiology. I propose that the process of circadian synchronization is analogous to protein folding dynamics, with energy minima across the circadian landscape achieved during phase alignment of individual cells and tissues (asynchronous oscillators) and misalignment (analogous to misfolding traps) induced by either environmental or behavioural perturbation. Entraining agents promote synchronization and circadian resonance of individual tissue clocks,

whereas circadian insults lead to off-synchrony pathways in which phase and amplitude are misaligned. Such misaligned states may be permanent (analogous to kinetic trapping of misfolded polypeptide) or re-aligned. I propose a concept of 'chronostasis' to describe the circadian synchrony landscape (just as proteostasis describes the folding landscape), with both *cis*- and *trans*-acting factors affecting achievement of energy minima and determining trajectory across the topological map.

(refs 72 and 73), delineating the relationship between circadian gene transcription changes and pathologies remains a challenge. The coincidence of circadian and metabolic perturbation in these mice may create a vicious cycle and augment the adverse effects of dual disruption in timing systems and metabolism.

Epigenetics and circadian cycles

Clocks also synchronize to the environment through post-translational modification of transcription factors and histones that tune gene expression rhythms to metabolic state (Fig. 5). NAD^+ oscillation, redox flux, ATP availability and mitochondrial function influence acetylation and methylation reactions, and may be important factors in circadian synchrony.

Chromatin transitions impact on core clock gene cycling

In liver tissue, the rhythmic acetylation of histone 3 corresponds with CLOCK–BMAL1-mediated transcriptional activation of *Per1* and *Per2* genes, and to recruitment of the histone acetyltransferase p300 — an event closely coupled to RNA polymerase II binding⁷⁷. In an additional twist, CLOCK itself participates in histone acetylation⁷⁸. Feedback repression by CRY proteins abrogates histone acetylation, which, together with rhythmic CLOCK–BMAL1 binding to DNA, probably contributes to circadian oscillation in gene expression⁷⁶. Further investigation will be necessary to determine how the rhythmic assembly of activator and repressor complexes influences the kinetics of transcriptional oscillation, and to delineate how metabolic signals modulate dynamic transitions in the epigenetic state.

Histone deacetylases couple metabolic and circadian cycles

Studies of histone deacetylase activity have demonstrated mechanistic integration between circadian and metabolic processes at the level of post-translational protein modification and gene transcription (Fig. 4). As already noted, REV-ERB functions together with the co-repressor NCOR1 to rhythmically recruit the class I histone deacetylase HDAC3, and genetic abrogation of the NCOR1–HDAC3 interaction results in both metabolic pathologies and circadian disruption⁷⁵. In addition, interactions occur between CLOCK–BMAL1 and the class III histone deacetylases belonging to the sirtuin (SIRT, silencer of information regulator) superfamily of chromatin-modifying enzymes^{79,80}. SIRT1 association with the clock activator complex represents an additional

mode of autoregulation because CLOCK–BMAL1 directly modulates the turnover of cellular NAD^+ , a cofactor for deacetylase reactions^{81,82}. NAD^+ functions as an electron shuttle in oxidoreductase reactions and as a cofactor in deacetylase and ADP-ribosylation modifications, raising intriguing questions concerning the role of NAD^+ in bidirectional interactions between circadian and metabolic signalling. Furthermore, the NAD^+ -dependent sirtuins have been established as regulators of metabolic pathways in response to calorie restriction and as modulators of oxidative damage and DNA-repair processes that are central to lifespan regulation. As such, the nexus of circadian control of NAD^+ and sirtuin activity may have broad implications for ageing and oxidative metabolism, which are particularly relevant in view of the association between period length and longevity⁸³.

Histone methylation in metabolism and circadian rhythms

In addition to identification of histone acetylation as a key event in circadian cycles, biochemical pull down with PER1 has demonstrated that the clock repressor complex is also associated with factors involved in histone methylation, including WDR5 (ref. 84), whereas the mammalian methyltransferase EZH2 has been shown to participate in CRY-mediated repression⁸⁵ (Fig. 4). Conversely, rhythmic recruitment of the histone methyltransferase MLL1 participates in gene activation by CLOCK–BMAL1 (ref. 86). Rhythmic trimethylation of histone 3 lysine 4 (H3K4me3) is also involved in circadian activation of the clock-controlling gene *DBP*, whereas dimethylation of histone 3 lysine 9 (H3K9me2) corresponds to its repression⁸⁷. The link between histone methylation and circadian cycles has also been identified in studies of the *Arabidopsis thaliana* clock⁸⁸. CLOCK–BMAL1 activity has also been shown to involve cyclic association of the histone demethylase JARID1a (ref. 89), although a direct link to methylation was not observed. Activity of the jumoni C domain demethylases, such as JARID1a, is coupled to cellular redox and mitochondrial energetics because both Fe II and α -ketoglutarate are used as cofactors⁹⁰. Whether changes in cellular metabolism may affect circadian systems through alterations in DNA methylation remains untested.

A range of additional mechanisms have been implicated in transduction of metabolic flux to co-activators and co-repressors, although the effect these may have on circadian gene transcription is not yet known (Fig. 4). For instance, CTBP1, a co-repressor associated with

NCOR1, is sensitive to the NADH:NAD⁺ ratio, and CTBP1 itself is an NAD⁺ reductase⁹¹, making this a potential candidate in the communication between metabolic changes and circadian oscillations.

Chronobiology and health and management of disease

The integrative physiology of circadian and metabolic systems has emerged through a combination of biochemical and experimental genetic studies; however, emerging approaches in human analyses provide an important avenue for future work (Fig. 5). Monogenic disorders in sleep onset and waking have provided evidence that clock genes have an effect not only on subjective chronotype, but also on neurological pathways that regulate sleep in humans. Animal models of these coding mutations in humans may be a platform with which to investigate links between neuroendocrine homeostasis and circadian systems⁹².

One of the more surprising results in human genetic analyses has been the association between variants of both *CRY2* and *MTNR1B* genes with glucose levels in humans^{93–95}. Although the pharmacology of the melatonin receptor 1B in glucose homeostasis is complex, studies in rodents support a role for melatonin signalling in rhythms of insulin secretion⁹⁶. Studies of Smith–Magenis syndrome, a haploinsufficiency disorder localized to *RAI1* and characterized by neural–behavioural abnormalities, intellectual deficit, obesity and circadian sleep disruption, has provided further evidence for a genetic link between *CLOCK* gene expression and energetic disorders⁹⁷. Emerging human genetic work parallels evidence from animal-based experimental studies to strengthen the hypothesis that genetic signatures of circadian function may be used to predict risk for metabolic disorders in humans.

Although beyond the scope of this Review, both longitudinal population studies and clinical investigations have indicated there is an association between shift work and metabolic disease. For instance, a study of nurses, who are one of the best monitored cohorts with a large representation of individuals who work shifts, has associated sleep time and circadian disruption with a broad range of disorders — including type 2 diabetes, gastrointestinal disorders and cancer — that may also be modulated by circadian genotype⁹⁸. Moreover, sleep loss and circadian disruption may be interacting risk factors for developing type 2 diabetes in individuals who are predisposed to the disease⁹⁹. The public health implications may be quite broad given the frequency of circadian behavioural disruption; indeed, the habit of altering bedtime on weekends, or ‘social jet lag’, has been associated with increased body weight¹⁰⁰. Laboratory models also suggest there is a direct causal role of circadian disruption on glucose tolerance⁴², although separating the effects of circadian disruption from sleep reduction as a result of experimental regimens remains a challenge. Ultimately, a combination of clinical, genetic and animal paradigms will be needed to understand the links between circadian biology and metabolism and to tailor preventive interventions and therapies for humans.

Future horizons and implications of time

Summer 2012 marked a four-year cycle when we celebrated breaking physical boundaries at the Olympic Games; we also saw NASA's Voyager 1 spacecraft reach the farthest distance a man-made object has journeyed from our planet. Despite the marvel of progress that these images conjure, in our realistic moments we are reminded of the primeval constraints of our simple origins on the surface of Earth, none more fundamental than the daily alternation of light and darkness. Indeed, recognition of this environmental pressure has marked thinking about evolution ever since Darwinian times, despite the abstraction of space travel. What is the meaning of this pervasive timescale?

In green plants, circadian cycles represent an innovation to the dominant constraint of time. Namely, sessile plants use clocks to defend against DNA damage during exposure to sunlight, while optimizing oxygenic photosynthesis. In animals, the circadian system also provides flexibility in response to environmental challenge, but the solutions involve adaptations within both the nervous system and peripheral tissues. Nonetheless, some of the same toolkit has been

deployed — including the PAS domain module, and the coupling of transcriptional oscillators to metabolic outputs. Epigenetic programs also fine-tune the clock across prokaryote and eukaryote lineages.

In the past century, we have also witnessed the invention of electric light, television, the jet engine and the Internet. But we are still unable to escape from the limits of inner time.

Although many gaps exist in our understanding, there is compelling evidence that points towards environmental disruptors of timing as agents of metabolic dysregulation (Fig. 5). Although abrogation through genetic mutation of the clock pathway is unlikely to explain common disease, the technology is now available to test how rare variants affect disorders such as type 2 diabetes, obesity and associated cardiometabolic complications. Clocks may inform stratification of risk of dyslipidaemia, microalbuminuria, retinopathy, neuropathy and cardiomyopathy. The discovery of circadian variants that affect glucose homeostasis, gluconeogenesis and β -cell function raises the possibility that pharmacological modification of molecular clock function may have therapeutic benefits. Given the window that circadian systems provide to understanding the partitioning and flux of fuel across different phases of the fasting–feeding cycle, it is likely that insight into the clock system may also provide an understanding of metabolic fate. This may include processes such as cellular regeneration and proliferation, and the switch from quiescent to active states in haematopoietic tissues. Although we are fixed in time by the clock in our genes, it is reasonable to predict that drugs, and even nutraceutical interventions, may soon be in hand to selectively alter time — even within specific tissues and cells — as a means of improving robustness, adaptability and health. ■

- Allada, R., Emery, P., Takahashi, J. S. & Rosbash, M. Stopping time: the genetics of fly and mouse circadian clocks. *Annu. Rev. Neurosci.* **24**, 1091–1119 (2001).
- Hardin, P. E., Hall, J. C. & Rosbash, M. Feedback of the *Drosophila* period gene product on circadian cycling of its messenger RNA levels. *Nature* **343**, 536–540 (1990).
- Loros, J. J. & Dunlap, J. C. *Neurospora crassa* clock-controlled genes are regulated at the level of transcription. *Mol. Cell. Biol.* **11**, 558–563 (1991).
- Hsu, D. S. *et al.* Putative human blue-light photoreceptors hCRY1 and hCRY2 are flavoproteins. *Biochemistry* **35**, 13871–13877 (1996).
- Kitayama, Y., Nishiwaki, T., Terauchi, K. & Kondo, T. Dual KaiC-based oscillations constitute the circadian system of cyanobacteria. *Genes Dev.* **22**, 1513–1521 (2008).
- Nakajima, M. *et al.* Reconstitution of circadian oscillation of cyanobacterial KaiC phosphorylation *in vitro*. *Science* **308**, 414–415 (2005).
This paper reports the reconstitution of the first circadian reaction *in vitro* in the presence of just protein and ATP.
- Rust, M. J., Golden, S. S. & O'Shea, E. K. Light-driven changes in energy metabolism directly entrain the cyanobacterial circadian oscillator. *Science* **331**, 220–223 (2011).
- Rutter, J., Reick, M., Wu, L. C. & McKnight, S. L. Regulation of clock and NPAS2 DNA binding by the redox state of NAD cofactors. *Science* **293**, 510–514 (2001).
This paper initiated the hypothesis that circadian cycles arise from metabolic cycles.
- O'Neill, J. S. & Reddy, A. B. Circadian clocks in human red blood cells. *Nature* **469**, 498–503 (2011).
- O'Neill, J. S. *et al.* Circadian rhythms persist without transcription in a eukaryote. *Nature* **469**, 554–558 (2011).
This work advanced the hypothesis that redox sensing occurs in eukaryotes independently of transcription and stimulate consideration of the origins of circadian oscillators.
- Edgar, R. S. *et al.* Peroxiredoxins are conserved markers of circadian rhythms. *Nature* **485**, 459–464 (2012).
- Dodd, A. N. *et al.* Plant circadian clocks increase photosynthesis, growth, survival, and competitive advantage. *Science* **309**, 630–633 (2005).
- Chen, Z., Odstrcil, E. A., Tu, B. P. & McKnight, S. L. Restriction of DNA replication to the reductive phase of the metabolic cycle protects genome integrity. *Science* **316**, 1916–1919 (2007).
- Provencio, I., Jiang, G., De Grip, W. J., Hayes, W. P. & Rollag, M. D. Melanopsin: an opsin in melanophores, brain, and eye. *Proc. Natl Acad. Sci. USA* **95**, 340–345 (1998).
- Guler, A. D. *et al.* Melanopsin cells are the principal conduits for rod–cone input to non-image-forming vision. *Nature* **453**, 102–105 (2008).
- Hattar, S., Liao, H. W., Takao, M., Berson, D. M. & Yau, K. W. Melanopsin-containing retinal ganglion cells: architecture, projections, and intrinsic photosensitivity. *Science* **295**, 1065–1070 (2002).
- Gooley, J. J., Lu, J., Chou, T. C., Scammell, T. E. & Saper, C. B. Melanopsin in cells of origin of the retinohypothalamic tract. *Nature Neurosci.* **4**, 1165 (2001).
- Stephan, F. K. & Zucker, I. Circadian rhythms in drinking behavior and locomotor activity of rats are eliminated by hypothalamic lesions. *Proc. Natl Acad. Sci. USA* **69**, 1583–1586 (1972).

19. Ralph, M. R., Foster, R. G., Davis, F. C. & Menaker, M. Transplanted suprachiasmatic nucleus determines circadian period. *Science* **247**, 975–978 (1990).
20. Chou, T. C. et al. Critical role of dorsomedial hypothalamic nucleus in a wide range of behavioral circadian rhythms. *J. Neurosci.* **23**, 10691–10702 (2003).
21. Landry, G. J. et al. Evidence for time-of-day dependent effect of neurotoxic dorsomedial hypothalamic lesions on food anticipatory circadian rhythms in rats. *PLoS ONE* **6**, e24187 (2011).
22. Sutton, G. M. et al. The melanocortin-3 receptor is required for entrainment to meal intake. *J. Neurosci.* **28**, 12946–12955 (2008).
23. Lin, L. et al. The sleep disorder canine narcolepsy is caused by a mutation in the hypocretin (orexin) receptor 2 gene. *Cell* **98**, 365–376 (1999).
This study placed the orexin pathway at the genetic intersection of sleep and metabolism.
24. Nishino, S. et al. Low cerebrospinal fluid hypocretin (orexin) and altered energy homeostasis in human narcolepsy. *Ann. Neurol.* **50**, 381–388 (2001).
25. Funato, H. et al. Enhanced orexin receptor-2 signaling prevents diet-induced obesity and improves leptin sensitivity. *Cell Metab.* **9**, 64–76 (2009).
26. Yamazaki, S. et al. Effects of aging on central and peripheral mammalian clocks. *Proc. Natl Acad. Sci. USA* **99**, 10801–10806 (2002).
27. Dubrovsky, Y. V., Samsa, W. E. & Kondratov, R. V. Deficiency of circadian protein CLOCK reduces lifespan and increases age-related cataract development in mice. *Aging (Albany NY)* **2**, 936–944 (2010).
28. Kondratov, R. V., Kondratova, A. A., Gorbacheva, V. Y., Vykhovanets, O. V. & Antoch, M. P. Early aging and age-related pathologies in mice deficient in BMAL1, the core component of the circadian clock. *Genes Dev.* **20**, 1868–1873 (2006).
29. Matsuo, T. et al. Control mechanism of the circadian clock for timing of cell division *in vivo*. *Science* **302**, 255–259 (2003).
30. Damiola, F. et al. Restricted feeding uncouples circadian oscillators in peripheral tissues from the central pacemaker in the suprachiasmatic nucleus. *Genes Dev.* **14**, 2950–2961 (2000).
The work in this report followed pioneering studies, by the same group, demonstrating cell-autonomous oscillation of the circadian clock in fibroblasts, and provided the first evidence for peripheral molecular clock entrainment to feeding.
31. Le Minh, N., Damiola, F., Tronche, F., Schutz, G. & Schibler, U. Glucocorticoid hormones inhibit food-induced phase-shifting of peripheral circadian oscillators. *EMBO J.* **20**, 7128–7136 (2001).
32. Buhr, E. D., Yoo, S. H. & Takahashi, J. S. Temperature as a universal resetting cue for mammalian circadian oscillators. *Science* **330**, 379–385 (2010).
33. Saini, C., Morf, J., Stratmann, M., Gos, P. & Schibler, U. Simulated body temperature rhythms reveal the phase-shifting behavior and plasticity of mammalian circadian oscillators. *Genes Dev.* **26**, 567–580 (2012).
34. Asher, G. et al. Poly(ADP-ribose) polymerase 1 participates in the phase entrainment of circadian clocks to feeding. *Cell* **142**, 943–953 (2010).
35. Kohsaka, A. et al. High-fat diet disrupts behavioral and molecular circadian rhythms in mice. *Cell Metab.* **6**, 414–421 (2007).
This study showed that a high-fat diet can perturb core properties of the internal clock.
36. Turek, F. W. et al. Obesity and metabolic syndrome in circadian *Clock* mutant mice. *Science* **308**, 1043–1045 (2005).
This article reports work that opened genetic approaches to probe links between clocks and metabolism.
37. Dudley, C. A. et al. Altered patterns of sleep and behavioral adaptability in NPAS2-deficient mice. *Science* **301**, 379–383 (2003).
38. Arble, D. M., Bass, J., Laposky, A. D., Vitaterna, M. H. & Turek, F. W. Circadian timing of food intake contributes to weight gain. *Obesity (Silver Spring)* **17**, 2100–2102 (2009).
39. Hatori, M. et al. Time-restricted feeding without reducing caloric intake prevents metabolic diseases in mice fed a high-fat diet. *Cell Metab.* **15**, 848–860 (2012).
40. Fonken, L. K. et al. Light at night increases body mass by shifting the time of food intake. *Proc. Natl Acad. Sci. USA* **107**, 18664–18669 (2010).
41. Spiegel, K., Leproult, R. & Van Cauter, E. Impact of sleep debt on metabolic and endocrine function. *Lancet* **354**, 1435–1439 (1999).
42. Scheer, F. A., Hilton, M. F., Mantzoros, C. S. & Shea, S. A. Adverse metabolic and cardiovascular consequences of circadian misalignment. *Proc. Natl Acad. Sci. USA* **106**, 4453–4458 (2009).
43. Shea, S. A., Hilton, M. F., Hu, K. & Scheer, F. A. Existence of an endogenous circadian blood pressure rhythm in humans that peaks in the evening. *Circ. Res.* **108**, 980–984 (2011).
44. Wang, J., Yin, L. & Lazar, M. A. The orphan nuclear receptor Rev-erb α regulates circadian expression of plasminogen activator inhibitor type 1. *J. Biol. Chem.* **281**, 33842–33848 (2006).
45. Schoenhard, J. A. et al. Regulation of the PAI-1 promoter by circadian clock components: differential activation by BMAL1 and BMAL2. *J. Mol. Cell. Cardiol.* **35**, 473–481 (2003).
46. Jeyaraj, D. et al. Circadian rhythms govern cardiac repolarization and arrhythmogenesis. *Nature* **483**, 96–99 (2012).
47. Bray, M. S. et al. Disruption of the circadian clock within the cardiomyocyte influences myocardial contractile function, metabolism, and gene expression. *Am. J. Physiol. Heart Circ. Physiol.* **294**, H1036–H1047 (2008).
48. Curtis, A. M. et al. Circadian variation of blood pressure and the vascular response to asynchronous stress. *Proc. Natl Acad. Sci. USA* **104**, 3450–3455 (2007).
49. Cheng, B. et al. Tissue-intrinsic dysfunction of circadian clock confers transplant arteriosclerosis. *Proc. Natl Acad. Sci. USA* **108**, 17147–17152 (2011).
50. Pan, X., Zhang, Y., Wang, L. & Hussain, M. M. Diurnal regulation of MTP and plasma triglyceride by CLOCK is mediated by SHP. *Cell Metab.* **12**, 174–186 (2010).
51. Douris, N. et al. Nocturnal regulates circadian trafficking of dietary lipid in intestinal enterocytes. *Curr. Biol.* **21**, 1347–1355 (2011).
52. Marcheva, B. et al. Disruption of the clock components CLOCK and BMAL1 leads to hypoinsulinaemia and diabetes. *Nature* **466**, 571–572 (2010).
53. Rudic, R. D. et al. BMAL1 and CLOCK, two essential components of the circadian clock, are involved in glucose homeostasis. *PLoS Biol.* **2**, e377 (2004).
54. Lamia, K. A., Storch, K. F. & Weitz, C. J. Physiological significance of a peripheral tissue circadian clock. *Proc. Natl Acad. Sci. USA* **105**, 15172–15177 (2008).
55. Zhang, E. E. et al. Cryptochrome mediates circadian regulation of cAMP signalling and hepatic gluconeogenesis. *Nature Med.* **16**, 1152–1156 (2010).
56. Lamia, K. A. et al. Cryptochromes mediate rhythmic repression of the glucocorticoid receptor. *Nature* **480**, 552–556 (2011).
57. Baggs, J. E. et al. Network features of the mammalian circadian clock. *PLoS Biol.* **7**, e52 (2009).
58. Dioum, E. M. et al. NPAS2: a gas-responsive transcription factor. *Science* **298**, 2385–2387 (2002).
59. Gilles-Gonzalez, M. A. & Gonzalez, G. Signal transduction by heme-containing PAS-domain proteins. *J. Appl. Physiol.* **96**, 774–783 (2004).
60. Marvin, K. A. et al. Nuclear receptors *Homo sapiens* Rev-erb β and *Drosophila melanogaster* E75 are thiolate-ligated heme proteins which undergo redox-mediated ligand switching and bind CO and NO. *Biochemistry* **48**, 7056–7071 (2009).
61. Yin, L. et al. Rev-erb α , a heme sensor that coordinates metabolic and circadian pathways. *Science* **318**, 1786–1789 (2007).
62. Raghuram, S. et al. Identification of heme as the ligand for the orphan nuclear receptors REV-ERB α and REV-ERB β . *Nature Struct. Mol. Biol.* **14**, 1207–1213 (2007).
63. Lamia, K. A. et al. AMPK regulates the circadian clock by cryptochrome phosphorylation and degradation. *Science* **326**, 437–440 (2009).
This study introduces a molecular mechanism for feedback regulation of the internal clock through metabolic flux.
64. Panda, S. et al. Coordinated transcription of key pathways in the mouse by the circadian clock. *Cell* **109**, 307–320 (2002).
This study applied genomic approaches to define the widespread circadian control of metabolic pathways.
65. Yang, X. et al. Nuclear receptor expression links the circadian clock to metabolism. *Cell* **126**, 801–810 (2006).
66. Dufour, C. R. et al. Genomic convergence among ERR α , PROX1, and BMAL1 in the control of metabolic clock outputs. *PLoS Genet.* **7**, e1002143 (2011).
67. Liu, C., Li, S., Liu, T., Borjigin, J. & Lin, J. D. Transcriptional coactivator PGC-1 α integrates the mammalian clock and energy metabolism. *Nature* **447**, 477–481 (2007).
68. Schmutz, I., Ripperger, J. A., Baeriswyl-Aebischer, S. & Albrecht, U. The mammalian clock component PERIOD2 coordinates circadian output by interaction with nuclear receptors. *Genes Dev.* **24**, 345–357 (2010).
69. Torra, I. P. et al. Circadian and glucocorticoid regulation of Rev-erb α expression in liver. *Endocrinology* **141**, 3799–3806 (2000).
70. Preitner, N. et al. The orphan nuclear receptor REV-ERB α controls circadian transcription within the positive limb of the mammalian circadian oscillator. *Cell* **110**, 251–260 (2002).
71. Solt, L. A. et al. Regulation of circadian behaviour and metabolism by synthetic REV-ERB agonists. *Nature* **485**, 62–68 (2012).
72. Bugge, A. et al. Rev-erb α and Rev-erb β coordinately protect the circadian clock and normal metabolic function. *Genes Dev.* **26**, 657–667 (2012).
73. Cho, H. et al. Regulation of circadian behaviour and metabolism by REV-ERB- α and REV-ERB- β . *Nature* **485**, 123–127 (2012).
74. Yin, L. & Lazar, M. A. The orphan nuclear receptor Rev-erb α recruits the N-CoR/histone deacetylase 3 corepressor to regulate the circadian *Bmal1* gene. *Mol. Endocrinol.* **19**, 1452–1459 (2005).
75. Feng, D. et al. A circadian rhythm orchestrated by histone deacetylase 3 controls hepatic lipid metabolism. *Science* **331**, 1315–1319 (2011).
This paper integrates genomic approaches to illustrate the epigenetic mechanisms that link circadian oscillation with metabolism.
76. Rey, G. et al. Genome-wide and phase-specific DNA-binding rhythms of BMAL1 control circadian output functions in mouse liver. *PLoS Biol.* **9**, e1000595 (2011).
77. Etchegaray, J. P., Lee, C., Wade, P. A. & Reppert, S. M. Rhythmic histone acetylation underlies transcription in the mammalian circadian clock. *Nature* **421**, 177–182 (2003).
78. Doi, M., Hirayama, J. & Sassone-Corsi, P. Circadian regulator CLOCK is a histone acetyltransferase. *Cell* **125**, 497–508 (2006).
79. Nakahata, Y. et al. The NAD $^{+}$ -dependent deacetylase SIRT1 modulates CLOCK-mediated chromatin remodeling and circadian control. *Cell* **134**, 329–340 (2008).
80. Asher, G. et al. SIRT1 regulates circadian clock gene expression through PER2 deacetylation. *Cell* **134**, 317–328 (2008).
References 79 and 80 report a link between the ageing related sirtuin deacetylases and circadian metabolism.
81. Ramsey, K. M. et al. Circadian clock feedback cycle through NAMPT-mediated NAD $^{+}$ biosynthesis. *Science* **324**, 651–654 (2009).
82. Nakahata, Y., Sahar, S., Astarita, G., Kaluzova, M. & Sassone-Corsi, P. Circadian control of the NAD $^{+}$ salvage pathway by CLOCK–SIRT1. *Science* **324**, 654–657 (2009).
References 81 and 82 make up work that defines a feedback loop linking NAD $^{+}$ biosynthesis to circadian oscillation.

83. Libert, S., Bonkowski, M. S., Pointer, K., Pletcher, S. D. & Guarente, L. Deviation of innate circadian period from 24 h reduces longevity in mice. *Aging Cell* **11**, 794–800 (2012).
84. Brown, S. A. *et al.* PERIOD1-associated proteins modulate the negative limb of the mammalian circadian oscillator. *Science* **308**, 693–696 (2005).
85. Etchegaray, J. P. *et al.* The polycomb group protein EZH2 is required for mammalian circadian clock function. *J. Biol. Chem.* **281**, 21209–21215 (2006).
86. Katada, S. & Sassone-Corsi, P. The histone methyltransferase MLL1 permits the oscillation of circadian gene expression. *Nature Struct. Mol. Biol.* **17**, 1414–1421 (2010).
87. Ripperger, J. A. & Schibler, U. Rhythmic CLOCK–BMAL1 binding to multiple E-box motifs drives circadian *Dbp* transcription and chromatin transitions. *Nature Genet.* **38**, 369–374 (2006).
88. Jones, M. A. *et al.* Jumoni domain protein JMJD5 functions in both the plant and human circadian systems. *Proc. Natl Acad. Sci. USA* **107**, 21623–21628 (2010).
89. DiTacchio, L. *et al.* Histone lysine demethylase JARID1a activates CLOCK–BMAL1 and influences the circadian clock. *Science* **333**, 1881–1885 (2011).
90. Tsukada, Y. *et al.* Histone demethylation by a family of JmjC domain-containing proteins. *Nature* **439**, 811–816 (2006).
91. Zhang, Q., Piston, D. W. & Goodman, R. H. Regulation of corepressor function by nuclear NADH. *Science* **295**, 1895–1897 (2002).
92. Xu, Y. *et al.* Modeling of a human circadian mutation yields insights into clock regulation by PER2. *Cell* **128**, 59–70 (2007).
93. Dupuis, J. *et al.* New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nature Genet.* **42**, 105–116 (2010).
94. Prokopenko, I. *et al.* Variants in *MTNR1B* influence fasting glucose levels. *Nature Genet.* **41**, 77–81 (2009).
95. Lyssenko, V. *et al.* Common variant in *MTNR1B* associated with increased risk of type 2 diabetes and impaired early insulin secretion. *Nature Genet.* **41**, 82–88 (2009).
96. Picinato, M. C., Haber, E. P., Carpinelli, A. R. & Cipolla-Neto, J. Daily rhythm of glucose-induced insulin secretion by isolated islets from intact and pinealectomized rat. *J. Pineal Res.* **33**, 172–177 (2002).
97. Williams, S. R., Zies, D., Mullegama, S. V., Grotewiel, M. S. & Elsea, S. H. Smith–Magenis syndrome results in disruption of *CLOCK* gene transcription and reveals an integral role for RAI1 in the maintenance of circadian rhythmicity. *Am. J. Hum. Genet.* **90**, 941–949 (2012).
98. Pan, A., Schernhammer, E. S., Sun, Q. & Hu, F. B. Rotating night shift work and risk of type 2 diabetes: two prospective cohort studies in women. *PLoS Med.* **8**, e1001141 (2011).
99. Knutson, K. L., Van Cauter, E., Zee, P., Liu, K. & Lauderdale, D. S. Cross-sectional associations between measures of sleep and markers of glucose metabolism among subjects with and without diabetes: the Coronary Artery Risk Development in Young Adults (CARDIA) Sleep Study. *Diabetes Care* **34**, 1171–1176 (2011).
100. Roenneberg, T., Allebrandt, K. V., Mew, M. & Vetter, C. Social jetlag and obesity. *Curr. Biol.* **22**, 939–943 (2012).

Acknowledgements I wish to thank G. Barish, K. Moynihan Ramsey and the anonymous reviewers for comments on the manuscript, as well as D. Levine and B. Marcheva for their help with the figures. I also thank my fellow time travellers, R. Allada, J. Takahashi and F. Turek, for their collegiality and discussions. Work towards this manuscript was supported by grants from the NIH Diabetes and Digestive and Kidney Diseases (R01DK090625), and Heart, Lung and Blood (R01HL097817) Institutes, National Institute on Aging (P01AG011412), the Chicago Biomedical Consortium Searle Funds, the American Diabetes Association (1-09-RA-07), the Juvenile Diabetes Research Foundation (1-2008-114) and the University of Chicago Diabetes Research and Training Center (P60 DK020595).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The author declares competing financial interests: details accompany the full-text HTML version of this paper at go.nature.com/ox45sv. Readers are welcome to comment on the online version of this article at go.nature.com/ox45sv. Correspondence should be addressed to J.B. (j-bass@northwestern.edu).

Central nervous system control of metabolism

Martin G. Myers Jr^{1,2,3} & David P. Olson⁴

Although it is a widely held thought that direct hormone action on peripheral tissues is sufficient to mediate the control of nutrient handling, the role of the central nervous system in certain aspects of metabolism has long been recognized. Furthermore, recent findings have suggested a more general role for the central nervous system in metabolic control, and have revealed the importance of a number of cues and hypothalamic circuits. The brain's contributions to metabolic control are more readily revealed and play a crucial part in catabolic states or in hormone deficiencies that mimic starvation.

The survival of multicellular organisms depends on the appropriate uptake and release of nutrients by major metabolic tissues. In the absence of continuous feeding, the availability of metabolic fuels (for example, glucose, fatty acids and amino acids) for use in tissues is maintained by storing nutrients, which are later released at the appropriate time and rate. Hormones that are secreted by the pancreatic islets of Langerhans modulate important aspects of nutrient uptake and storage (insulin) or their release into the circulation (glucagon). They do this partly by acting directly on the tissues that are the main reservoirs for these nutrients (for example, liver, adipose and muscle tissue).

Although the direct actions of these hormones on metabolic tissues are crucial for whole-body metabolic homeostasis, several central nervous system (CNS)-regulated systems have long been recognized to control important aspects of metabolism. For instance, the elaboration of catecholamines by the sympathetic nervous system (SNS), along with hypothalamically controlled hormones (such as glucocorticoids and thyroid hormone), functions in concert with glucagon to mediate the counter-regulatory response, promoting nutrient release into the blood and favouring substrate use over storage¹. Well-recognized CNS pathways mediate crucial aspects of this counter-regulatory response. Recent data reveal a more general role for CNS pathways in the modulation of metabolism: various nutrient, energetic and hormonal cues (such as insulin and the adipose-derived hormone leptin) function in the hypothalamus to control glucose and lipid metabolism, in addition to overall energy balance^{2,3} (Fig. 1). Several hypothalamic areas, especially the ventromedial nucleus and arcuate nucleus (including the arcuate melanocortin system), as well as the hindbrain, make important contributions to these effects. Although the peripheral response to the hormones that control metabolism (for example, insulin and glucagon) tends to obscure these CNS contributions under many conditions, the brain controls overall metabolic tone and is crucial when the peripheral systems cannot compensate — especially when pancreatic hormone output is absent or cannot be modulated.

Established autonomic and neuroendocrine roles

As anyone who has examined parameters of glucose homeostasis in mammals knows, stresses (such as pain or restraint) increase blood sugar. This reflects the activation of the SNS and the hypothalamic–pituitary–adrenal axis, which promote the breakdown of macromolecular

forms of stored energy (such as glycogen, triglyceride and protein) for release into the bloodstream (as glucose, fatty acids and glycerol, and amino acids). Furthermore, SNS-mediated catecholamine action in the islets promotes increased glucagon release while suppressing insulin secretion; the parasympathetic nervous system (PNS) mediates opposing actions⁴. Therefore, stress signals, which are coordinated by the CNS, modulate autonomic and neuroendocrine function to increase blood glucose. Furthermore, this stress-induced hyperglycaemia complicates the treatment of diabetes by promoting poor glycaemic control⁵.

CNS-mediated modulation of glucose homeostasis is not unique to anxiogenic stresses, but it is also a crucial response to physiological stresses, such as exercise or infection, in which increased circulating nutrient availability is required to meet the needs of muscle contraction or the immune response, respectively⁶. Indeed, metabolic cues can also promote this response: hypoglycaemia — a common manifestation of iatrogenic insulin overdose — not only promotes glucagon secretion over insulin production at the level of the islets, but also acts directly in the brain to increase SNS activity and glucocorticoid production (as well as food intake), directing the return to normoglycaemia^{7–9}. Within the brain, specialized glucose-sensing neurons that lie in several sites in the brainstem (including the nucleus tractus solitarius and other parts of the dorsal vagal complex) and hypothalamus (such as the ventromedial nucleus) respond to the decreased availability of glucose (the main metabolic fuel for the brain under most conditions) to promote this counter-regulatory response⁹. The importance of this system is underscored by the consequences of its failure to respond appropriately to hypoglycaemia in individuals, following repeated exposure to hypoglycaemia as a consequence of intensive insulin therapy. Such impairment increases the severity and frequency of hypoglycaemic events, and diminishes the patient's ability to effectively recognize their occurrence.

A larger role for the CNS in metabolic control

In addition to the well-established role for the brain in mediating important aspects of the counter-regulatory response to hypoglycaemia and other stresses, results over the past 15 years have demonstrated a more general role for the CNS in the sensing and control of whole-body metabolic homeostasis^{3,10}. This theme has emerged from the search for signals that communicate the status of adiposity or energy stores to the brain to control food intake and energy

¹Division of Metabolism, Endocrinology and Diabetes, Department of Internal Medicine, University of Michigan, Ann Arbor, Michigan 48105, USA; ²Department of Molecular and Integrative Physiology, University of Michigan, Ann Arbor, Michigan 48105, USA; ³Neuroscience Program, University of Michigan, Ann Arbor, Michigan 48105, USA; and ⁴Division of Endocrinology, Department of Pediatrics, University of Michigan, Ann Arbor, Michigan 48105, USA.

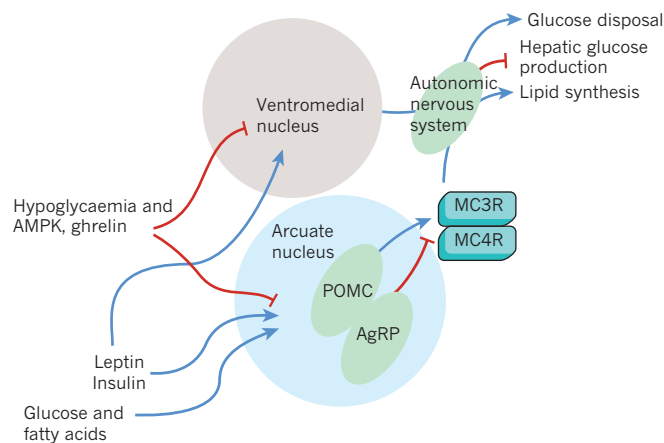


Figure 1 | Hypothalamic pathways in the control of metabolism. Recent findings have implicated medial basal hypothalamic nuclei (the arcuate and ventromedial nucleus) in the control of metabolism. Leptin, secreted from adipose tissue as an indicator of long-term energy stores, and insulin, reflective of recent food intake in addition to adipose stores, function in these nuclei to modulate the autonomic nervous system to increase hepatic glucose production and peripheral glucose disposal. Other signals of energy surfeit, including glucose and fatty acids, act in concert with these signals, whereas signals of energy deficit (such as hypoglycaemia and activation of AMP-dependent protein kinase (AMPK) and ghrelin) act in the opposite manner. Although the ventromedial nucleus neural pathways that mediate the responses to these stimuli have not been molecularly characterized, the arcuate nucleus melanocortin system (which is composed of pro-opiomelanocortin (POMC) neurons that act in part through melanocortin-3 receptor (MC3R) and melanocortin-4 receptor (MC4R), and is antagonized by Agouti-related protein (AgRP)-producing neurons) contributes.

expenditure. Insulin — the first hormone to be studied in this regard — exhibits elevated baseline blood concentrations with increasing adiposity, and increases rapidly in response to feeding. Acute administration of insulin directly into the CNS tends to suppress feeding¹⁰.

The cloning of the gene that encodes leptin, which is secreted by adipose tissue in approximate proportion to body fat stores, was a watershed event in our understanding of the role of the CNS in the control of energy balance and metabolism¹¹. A lack of leptin or its predominantly CNS-expressed receptor, LRB, not only promotes massive hyperphagia and decreased energy expenditure (with subsequent obesity), but also results in early-onset insulin resistance, hyperglycaemia and metabolic dysfunction, which are, at least in rodents, disproportionately severe in relation to that expected with obesity alone. Indeed, whole-body or CNS-restricted leptin treatment of leptin-deficient *ob/ob* mice rapidly restores glycaemic control independently of changes in food intake or adiposity, suggesting that leptin acts in the brain to control blood glucose levels independently of energy balance^{12,13}. These observations provoked a careful examination of the roles for CNS insulin and leptin action in the control of whole-body metabolism (mainly the control of glucose production and disposal)^{13,14}. These studies reported that CNS insulin action contributes to the suppression of hepatic glucose output, and CNS leptin action promotes increased hepatic glucose flux by increasing gluconeogenesis while diminishing glycogenolysis. In diet-induced obese rodents, leptin mainly suppresses glycogenolysis to decrease net hepatic glucose production (HGP)^{15,16}. In a similar vein, adiponectin and glucagon-like peptide 1 (GLP-1), like insulin, suppress gluconeogenesis and HGP following intracerebroventricular administration (although GLP-1-producing CNS neurons, rather than the gut, could represent the physiological source of CNS GLP-1 in this case)^{17,18}. By contrast, intracerebroventricular ghrelin (which opposes leptin action in most cases) increases HGP, as does the adipose-derived hormone resistin (which attenuates insulin action)^{19,20}.

Subsequent genetic studies have demonstrated crucial roles for CNS leptin and insulin signals (including those mediated by elements of

the insulin receptor substrate (IRS) 2 phosphatidylinositol (PI)-3-OH kinase–phosphoinositide-dependent kinase 1 (Pdk1) pathway that suppress the transcription factor FOXO1) in the control of glucose homeostasis^{21,22} (Fig. 2). Furthermore, impairing leptin and insulin action in the same CNS circuits provokes substantially greater metabolic dysfunction than that observed with ablation of either one alone, suggesting a synergy between the insulin and leptin signalling pathways in the CNS control of metabolism²³ (Fig. 2).

In addition to hormones, nutrients and signals of cellular energy status also have a role in the CNS control of metabolism. As might be predicted from the response to hypoglycaemia, pathways that monitor cellular energy status in the hypothalamus also modulate food intake and whole-body metabolism. Activation of hypothalamic AMP-dependent protein kinase (AMPK) (which is stimulated by the depletion of cellular ATP levels) increases food intake and circulating glucose concentrations^{24,25}. Roles in the CNS for amino acids, their sensing by the mammalian target of rapamycin pathway and mitochondria-derived reactive oxygen species in the control of peripheral metabolism have not yet been examined. However, these cellular signalling systems contribute to the control of food intake and energy homeostasis^{26,27}, and could also participate in the CNS control of metabolism.

In addition to the response to hypoglycaemia, glucose sensing in the hypothalamus also controls peripheral glucose handling: intracerebroventricular glucose administration decreases HGP through a mechanism that requires cellular ATP generation and the subsequent closure of ATP-sensitive potassium (K^+ -ATP) channels³. Similarly, although fatty acids are not typically thought of as a major fuel for the CNS, in the hypothalamus fatty acids and the systems that mediate their mitochondrial import and oxidation control food intake and blood glucose levels²⁸. Intracerebroventricular injection of oleic acid (and, to a lesser extent, other fatty acids) suppresses HGP, as do a variety of pharmacological and short-hairpin-RNA-mediated manipulations designed to alter hypothalamic fatty-acid oxidation and increase cellular concentrations of fatty-acid-coenzyme A (CoA). These studies suggest there is an important role for fatty-acid metabolism, and potentially for fatty-acid-CoA as an intracellular mediator, in nutrient sensing by hypothalamic neurons that modulate hepatic glucose handling.

Although most data that link the CNS to peripheral metabolism have examined the control of blood glucose concentrations as their surrogate for metabolism, a number of studies also suggest that these pathways may similarly modulate lipid handling throughout the body. Intracerebroventricular injection of leptin, for instance, suppresses lipogenesis in adipose tissue²⁹. Moreover, genetic or pharmacological manipulation of CNS melanocortin action can alter systemic lipid metabolism³⁰.

CNS circuits and cellular signals in metabolic control

Brain-wide and circuit-specific genetic manipulation of intracellular mediators of leptin or insulin action have demonstrated important roles for these signals in the CNS for the control of metabolism. Although genetic tools that manipulate many components of the cellular nutrient-sensing apparatus are not available or practical (because of factors such as redundant isoforms or embryonic lethality), direct injection of pharmacological inhibitors or short-hairpin RNA reagents to target these pathways has suggested that the hypothalamus, especially the medial basal hypothalamus (MBH) — which includes the ventromedial and arcuate nuclei — is an important site of action³. Other hypothalamic regions (such as the lateral hypothalamic area) and the brainstem contain neural populations that sense leptin and nutrients, including glucose and amino acids, but roles for these areas in the control of metabolism by direct leptin or insulin action have not been reported^{9,31}. However, the dorsal vagal complex (including the nucleus tractus solitarius) participates in the sensing of hypoglycaemia and the counter-regulatory response, and represents a potential contributor to overall metabolic

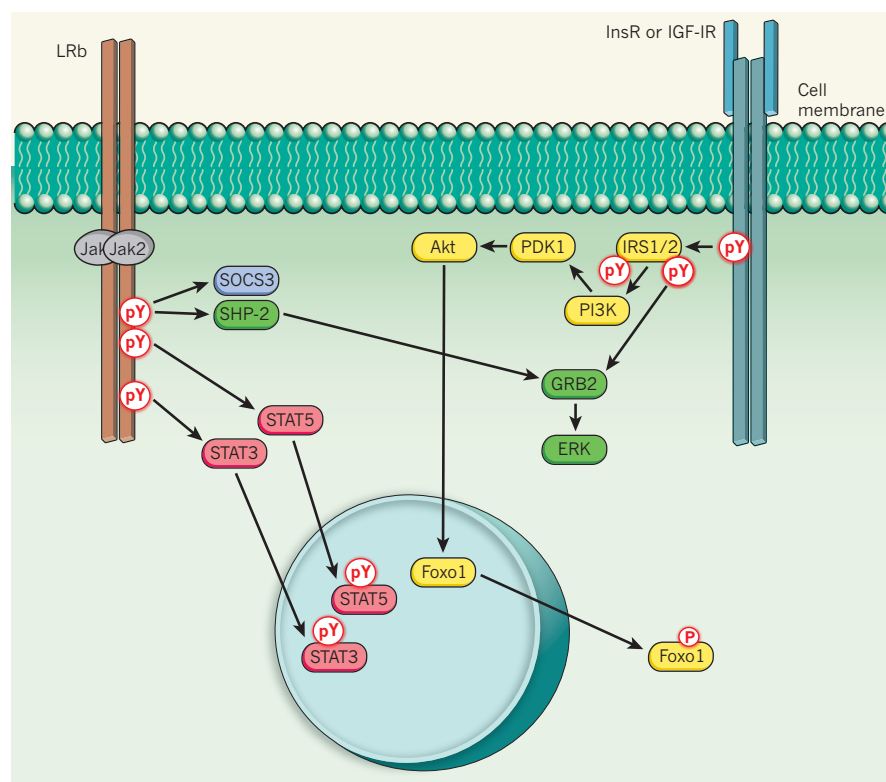


Figure 2 | Cellular signalling pathways modulated by leptin and insulin. The insulin receptor (InsR) and the related insulin-like growth factor 1 receptor (IGF-1R) act through their intrinsic tyrosine kinases to promote the phosphorylation of receptor tyrosine residues (pY), leading to the recruitment and phosphorylation of the insulin receptor substrate (IRS) proteins-1 and -2. These recruit phosphatidylinositol 3-kinase (PI3K), which activates phosphoinositide-dependent kinase 1 (PDK1) and Akt to promote, among other things, the phosphorylation and nuclear exclusion of the FOXO1 transcription factor, inactivating FOXO1-mediated transcription. Phosphorylated IRS1 or IRS2 also recruit GRB2, which promotes extracellular-signal-regulated kinase (ERK) activation. Stimulation of leptin receptor (LRb) activates the associated Jak2 tyrosine kinase to promote the phosphorylation of intracellular tyrosine residues on LRb. One of these residues recruits the suppressor of cytokine signalling 3 (SOCS3) and the protein-tyrosine phosphatase SHP-2. SHP-2 recruits GRB2 to activate ERK signalling. Two additional phosphorylated LRb residues recruit latent signal transducers and activators of transcription (STAT3 and STAT5), which then translocate to the nucleus to modulate gene transcription. LRb also acts through undefined pathways to modestly promote PI3K pathway signalling.

control. The dorsal motor nucleus of the vagus controls outputs to the periphery through the vagus nerve, and it is well-positioned to regulate metabolism.

The heterogeneity of brain tissue — in which even small and circumscribed brain areas (such as the arcuate nucleus) contain multiple types of neurons with distinct, or even opposing, functions — means that understanding the specific cell types and the circuits that are involved in metabolic control requires the use of cell-specific genetic tools^{32,33}. Although it is difficult to analyse the cellular pathways that control fatty-acid-CoA levels and sensing of cellular energy using neuron-specific techniques, many aspects of insulin and leptin action can be investigated with the tools that are available. For instance, given the important and synergistic roles that insulin and leptin have in the CNS control of metabolism, insulin signalling in LRb neurons might be an important mechanism to analyse. The redundancy of signalling by insulin and insulin-like growth factor 1 (IGF-1) receptors renders it difficult to analyse through the simple deletion of insulin receptors from LRb neurons (indeed, the metabolic phenotype of insulin-receptor deletion throughout the brain is rather modest). However, deletion of IRS2 — which is the crucial shared second messenger protein for insulin and IGF-1 receptors — from the entire brain or specifically from LRb neurons results in early insulin resistance and glucose intolerance^{34–35}.

Additionally, roles for direct LRb signalling in the control of energy homeostasis have been examined in mouse models, in which LRb mutants that are defective for specific signalling pathways replace the endogenous LRb. This analysis revealed crucial roles for LRb–signal transducers and activators of transcription 3 (STAT3) signalling in the control of energy balance, although the mice demonstrated improvements in glucose homeostasis relative to mice that were completely null for LRb^{36–38}. However, the improvements in glycaemic control may be secondary to increased overall glucose disposal because HGP remains high in animals that carry mutations that block LRb–STAT3 signalling. Another pathway, potentially mediated by the LRb-associated tyrosine kinase Jak2, or other LRb signals that are distinct from phosphorylation sites on LRb, also seems to contribute to the control of glucose homeostasis by leptin^{39,40}.

Thus, specific leptin- and insulin-mediated signals in LRb neurons contribute to glucose homeostasis. LRb neurons may be an important locus at which cellular nutrient and energy sensing could also have a notable metabolic role, although this has yet to be tested.

A role for the ventromedial nucleus

Within the MBH, the ventromedial nucleus is an obvious place to begin examining the hypothalamic circuits that control overall metabolism, owing to this region's important role in sensing glucose levels and in aspects of the counter-regulatory response^{41–43}. Manipulation of AMPK activity in the ventromedial nucleus by stereotactic injection of inhibitors or activators suggests a role for AMPK signalling in this nucleus in sensing low glucose concentrations and mounting the counter-regulatory response to hypoglycaemia. A role for the ventromedial nucleus in glucose metabolism has also been implied by the manipulation of glucokinase and the K⁺–ATP channel by stereotactic injection of pharmacological agents or short-hairpin RNA reagents into this nucleus^{43,44}.

Leptin activates at least some neurons in the ventromedial nucleus, which contains a substantial group of LRb neurons⁴⁵. This region also seems to have a role in the control of glucose disposal by leptin. The injection of leptin into the ventromedial nucleus, but not the adjacent arcuate nucleus, increases sympathetic output and glucose uptake by muscle and brown adipose tissue (both of which have heavy SNS innervation)⁴⁶. The deletion of LRb specifically from the ventromedial nucleus in mice has shown a role for ventromedial-nucleus leptin action in the control of food intake and overall energy expenditure^{47,48}. Although glucose metabolism in these animals has not been completely characterized, they are less glucose tolerant at early ages than control mice — before the onset of frank obesity — consistent with the idea that leptin action in the ventromedial nucleus promotes glucose disposal independently of the control of energy balance.

Insulin action in the ventromedial nucleus also contributes to the control of metabolism. Injection of insulin into this area promotes peripheral glucose disposal, whereas knockdown of insulin receptors in the ventromedial nucleus of adult mice causes glucose intolerance⁴⁹. Similarly, enhancing the action of insulin in the ventromedial nucleus by interference of the insulin-inhibited FOXO1 pathway

promotes increased glucose disposal in muscle tissue⁵⁰.

Although local hypoglycaemia (or AMPK activation) in the ventromedial-nucleus promotes a counter-regulatory response and glucose production, ventromedial-nucleus leptin and insulin signals promote glucose uptake by metabolically active tissues. Thus, whereas hypoglycaemia, and leptin or insulin in the ventromedial nucleus promote SNS-mediated processes, the specific outcome differs between stimuli. This discrepancy could reflect differences in the effects of each stimulus on the same cell type, or it might indicate that distinct (ventromedial nucleus) cell types respond to each stimulus. Unfortunately, systems to molecularly define and genetically perturb subtypes of ventromedial-nucleus-restricted neurons still need to be developed.

Arcuate nucleus and hypothalamic melanocortin system

Neural systems in the arcuate nucleus are thought to have a crucial role in the control of glucose homeostasis. Ablation of all arcuate-nucleus neurons with monosodium glutamate produces not only obesity but also frank hyperglycaemia, suggesting that the net tone of the arcuate nucleus tends to suppress hepatic glucose output⁵¹. By contrast, ventromedial hypothalamus lesions produce obesity, but also impair hypoglycaemia sensing. Of note, however, is that lesioning studies reveal nothing about the roles of specific populations of neurons within these nuclei, but only reveal the aggregate output for the entire area.

Arcuate-nucleus-directed LRb expression in rats and mice that are null for LRb alters body weight and food intake only modestly, but markedly ameliorates hyperglycaemia in these animals^{52,53}. Similarly, adenoviral augmentation of insulin signalling pathways in the arcuate nucleus also improves glucose homeostasis in LRb-null animals⁵⁴. Thus, leptin and insulin-like signalling in the arcuate nucleus are important for the suppression of glucose production.

Our understanding of the molecular make-up of neurons in the arcuate nucleus has advanced markedly over the past 15 years^{33,55}. This advance has permitted the development of a substantial array of molecular genetic tools with which to probe the function of specific sets of arcuate-nucleus neurons. Although the arcuate nucleus contains other cell types, two main opposing types of output neurons have been defined: those that express pro-opiomelanocortin and those that contain Agouti-related protein (AgRP) and neuropeptide-Y. In the arcuate nucleus, pro-opiomelanocortin is processed to produce α -melanocyte stimulating factor (α -MSH) — an agonist for the melanocortin-3 receptor (MC3R) and melanocortin-4 receptor (MC4R) — which collectively promote activity, energy expenditure and suppress food intake. AgRP is an inverse agonist for MC3R and MC4R, opposing central melanocortin action, and neuropeptide-Y is an inhibitory neuropeptide that suppresses energy use and promotes food intake. AgRP neurons also contain and release the inhibitory neurotransmitter GABA (γ -aminobutyric acid) as do most of the other non-pro-opiomelanocortin neurons in the arcuate nucleus⁵⁶. Insulin and leptin both inhibit AgRP neurons, and insulin action on these neurons suppresses HGP^{57,58}.

Pro-opiomelanocortin neurons and their roles in controlling glucose homeostasis are more complicated. Recent data suggest there are multiple subtypes of pro-opiomelanocortin neurons, including those that respond to leptin or serotonin (each of which activates distinct subtypes of pro-opiomelanocortin neurons) and those that are inhibited by insulin²³. Additional arcuate-nucleus neurons that transiently express pro-opiomelanocortin during development, but that later switch their fate to non-pro-opiomelanocortin-expressing cells have also been identified⁵⁹. Pharmacological studies suggest that melanocortin action increases hepatic glucose output, as does insulin action on pro-opiomelanocortin neurons in a background null for insulin receptors outside the liver^{15,58}. By contrast, expression of LRb in pro-opiomelanocortin neurons in an otherwise LRb-null background suppresses HGP, whereas simultaneous deletion of LRb and insulin receptors from pro-opiomelanocortin neurons increases HGP^{57,60,61}.

Impairing PI(3)K action in all pro-opiomelanocortin neurons also promotes insulin resistance⁶². Thus, although it is clear that the action of insulin and leptin on pro-opiomelanocortin neurons — along with the action of melanocortin in general — modulates HGP, the exact role of this pathway remains unclear. Presumably, the exact genetic, hormonal and physiological milieu of the animal affects the response to melanocortin activation. Although the relevant variables remain to be defined, the indirect regulation of pro-opiomelanocortin neurons by leptin action on other hypothalamic neurons could be a contributor⁵⁶.

Genetic or pharmacological manipulation of central melanocortin action also alters systemic lipid metabolism in rodents³⁰. Specifically, inhibition of melanocortin action promotes lipid uptake, triglyceride synthesis and accumulation of fat in white adipose tissue; central melanocortin blockade also increases circulating high-density lipoprotein cholesterol⁶³. Importantly, these effects are independent of food intake and body weight. Central melanocortin stimulation results in fat-depot-specific increases in phosphorylated perilipin A and hormone-sensitive lipase, presumably through sympathetic outflow⁶⁴. Whether these effects of melanocortin action will be recapitulated in humans remains to be determined, but these findings raise the possibility that central melanocortin receptors may be a therapeutic target for the lipid disorders associated with obesity and diabetes.

Melanocortin output

MC3R and MC4R are the predominant melanocortin receptors in the CNS. MC4R is expressed fairly widely in the CNS, whereas MC3R has more limited expression. Both receptors are expressed in the hypothalamus: MC3R is expressed heavily in the arcuate nucleus and ventromedial nucleus, whereas MC4R is expressed particularly strongly in the paraventricular hypothalamic nucleus.

Whole-body deletion of MC4R results in massive hyperphagic obesity, increased lean mass, and hyperinsulinaemia that is out of proportion to the obesity⁶⁵. Much of the metabolic disturbance observed in *Mc4r*-knockout mice may be secondary to their obesity, although *Mc4r*-knockout mice display decreased energy expenditure and increased metabolic efficiency prior to the onset of hyperphagia⁶⁶. *Mc3r*-knockout mice on the other hand do not display increased food intake and are more modestly obese, but they also have decreased lean mass and an abnormality in fuel partitioning that preferentially stores energy as fat⁶⁷. When challenged with a high-fat diet, male *Mc3r*-knockout mice gain weight rapidly and develop hyperinsulinism similar to that seen with *Mc4r*-knockout mice^{67–69}. This suggests a potentially important role for MC3R in mediating metabolic (rather than energy balance) responses to melanocortin action, although the extent to which this reflects altered nutrient partitioning relative to direct control of other metabolic parameters remains unclear. The function of these two melanocortin receptors in energy balance does not overlap because deletion of both receptors results in greater obesity and metabolic dysfunction than does the loss of either receptor alone⁶⁹. This additive effect underscores the importance of delineating the identity and neural circuitry used by both MC3R and MC4R⁷⁰.

Unfortunately, the low expression of MC3R and MC4R in most melanocortin-responsive neurons has prevented the detailed mapping of the melanocortin circuitry with much cellular specificity. Nevertheless, pharmacological or viral manipulation of melanocortin action by site-specific injection in multiple areas of the brain has been shown to alter energy balance^{71,72}. The effects of such manoeuvres on the metabolic actions of central melanocortins have not been fully explored. Genetic approaches have also been used to identify the specific neurons and circuits under melanocortin control that regulate metabolism, and these have suggested that central melanocortin action is dissociable in the CNS, with discrete brain regions mediating specific melanocortin effects (Box 1 and Fig. 3). Whether the metabolic effects of melanocortins are mediated mainly through neuroendocrine pathways or the autonomic nervous system or both has yet to be definitively determined.

BOX 1

Sites of melanocortin action

Genetic and pharmacological studies have begun to map the functional effects of central melanocortin action on specific regions of the brain. Pharmacological approaches using melanocortin agonists and antagonists have demonstrated a role for melanocortin action in the hindbrain and central amygdala in the regulation of food intake^{78–80}. Stereotactic injection of viral vectors that express melanocortin-receptor modulators has also suggested a role for the paraventricular nucleus, ventromedial nucleus and lateral hypothalamic area in mediating the effects of melanocortin action on food intake⁷². Such studies are unable to assign these physiological responses to a specific melanocortin receptor, because commonly used melanocortin agonists engage both MC3R and MC4R. Genetic approaches using neuron-specific Cre–*loxP* technologies have begun to address this issue^{81,82}: selective re-expression of MC4R in an otherwise null background has revealed the importance of neurons

within the paraventricular nucleus (and possibly the amygdala) in mediating the anorectic actions of MC4R. The role of these neurons in directly controlling metabolic effects is hard to decipher given the profound affect that restoring MC4R on these neurons — in an otherwise MC4R-null background — has on the obesity phenotype. By contrast, selective expression of MC4R in cholinergic, preganglionic autonomic output neurons alters energy expenditure and hepatic-insulin sensitivity in the obese *Mc4r*-knockout background. It is important to note that the subcellular location of functional MC4R (cell body compared with axon terminal) has not been clearly established, and it is possible that melanocortin ligands may exert their effects on melanocortin receptors expressed at terminals at a distance from the cell body. The development of similar genetic tools for MC3R should permit a more detailed neuron- and receptor-specific analysis of central melanocortin action⁸³.

The brain and the physiological control of metabolism

Although the participation of the brain in the response to hypoglycaemia is well-established and has an impact on the therapy for patients with insulin-deficient diabetes, the more recently described role of sensing of hormones and nutrients in the brain for the control of whole-body metabolism was initially met with some scepticism⁷³. Specifically, the roles of these brain systems have been defined largely at the limits of physiology — in many cases, under hyperinsulinaemic ‘clamp’ conditions that lock the contribution of the islets at a specific level, so that alterations in insulin or glucagon

secretion cannot compensate for changes in glucose production or disposal. The study of these brain systems often requires such conditions because the reserve insulin and glucagon secretory capacity of the islets can otherwise compensate for alterations in glucose handling by the major metabolic organs, obscuring the effect of the brain on the control of glucose production and disposal. To what extent these brain systems are relevant is controversial. They do, however, have a role in leptin deficiency: not only does leptin ameliorate hyperglycaemia in animals³ that are genetically leptin deficient, but it also improves metabolic function — by decreasing

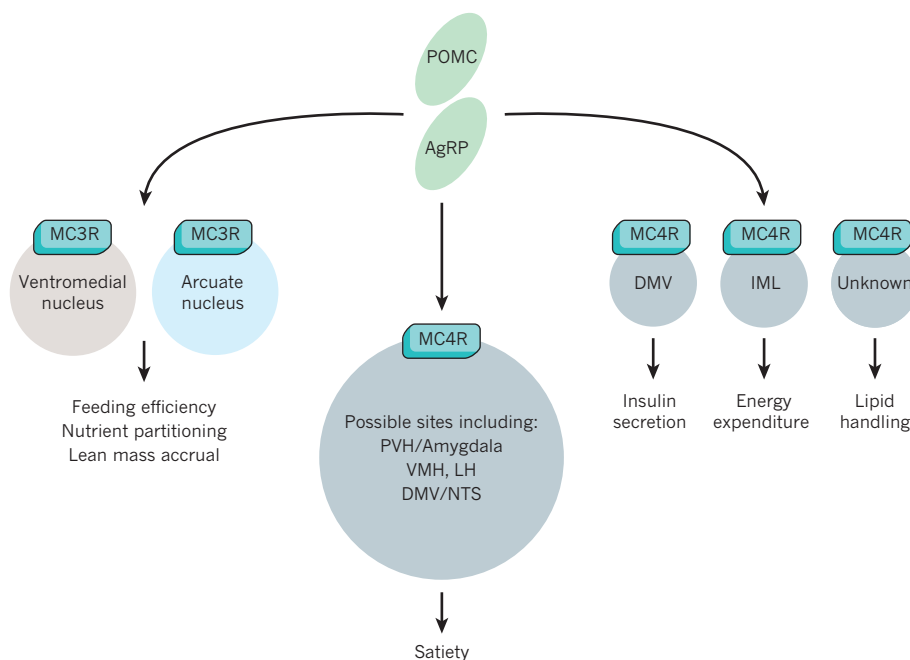


Figure 3 | CNS melanocortin action. The pro-opiomelanocortin (POMC) neurons of the arcuate nucleus produce melanocortin-receptor agonists, whereas Agouti-related protein (AgRP)-producing neurons antagonize melanocortin action. The two predominant CNS melanocortin receptors, MC3R and MC4R, mediate distinct effects. MC3R is expressed at high levels in the arcuate nucleus and ventromedial nucleus, and mainly controls the conversion of food to fat, nutrient partitioning and build-up of lean mass, whereas MC4R predominantly mediates effects on food intake. The specific neurons that express MC4R are unclear; however, genetic data suggest important roles for the paraventricular hypothalamic (PVH)

nucleus and amygdala MC4R in these effects, whereas other data, including pharmacological data, suggest roles for MC4R at other sites such as the ventromedial (VMH) nucleus, lateral hypothalamus (LH), dorsal motor nucleus of the vagus (DMV) and nucleus tractus solitarius (NTS), as well as the lateral parabrachial nucleus. In addition to mediating metabolic effects attributable to body weight, MC4R mediates the control of insulin secretion through the dorsal motor nucleus of the vagus and energy expenditure through the intermediolateral cell column (IML) of the spinal cord. MC4R also contributes to lipid handling in the body, but the MC4R-expressing site or sites that mediate these effects remain undefined.

blood sugar and lipids — in humans and mice with lipodystrophy syndromes^{74,75}. In lipodystrophy (in which the lack of adipose tissue results in functional leptin deficiency), circulating insulin concentrations are elevated, and exogenous insulin administration fails to adequately suppress hyperglycaemia and hyperlipidaemia. Thus, in this situation, leptin action (presumably in the brain) improves metabolic parameters that cannot be ameliorated by very high insulin levels. More commonly, brain leptin and insulin make important contributions to metabolic homeostasis in β -cell failure (for example, in insulin-deficient diabetes)^{76,77}. In rodents with insulin insufficiency due to chemical or autoimmune β -cell destruction, leptin (including intracerebroventricular leptin) normalizes blood glucose in the absence of exogenous insulin, and improves glycaemic control as an adjunct to insulin therapy. Leptin action is also likely to have an important role in glucoprivation, in which insulin is either fixed and artificially high (in insulin overdose) or fixed near zero (with inhibitors of glucose metabolism, such as 2-deoxyglucose).

Thus, the need to eliminate the islet contribution to metabolic control in order to observe the effects of the brain-hormone and nutrient-sensing pathways does not diminish the importance of these brain systems in the control of peripheral metabolic homeostasis. Instead, it suggests that the CNS serves to raise and lower the overall tone of the peripheral response, and dominates the control of metabolic homeostasis when pancreatic islets are unable to compensate.

Questions for the future

Although we now have a much clearer picture about how the brain contributes to metabolic homeostasis than we did a decade ago, many questions still remain. CNS perturbations have been seen to modulate processes such as glucose production and lipid metabolism in the periphery, but the circuits and mechanisms that intervene between the brain and systemic nutrient handling are poorly defined. Certainly, the SNS contributes to some of the outcomes, such as in the promotion of glucose disposal into skeletal muscle. Similarly, vagotomy — including sub-diaphragmatic vagotomy — prevents many of the changes in hepatic-glucose metabolism that occur in response to intracerebroventricular or MBH-directed injections, suggesting that the parasympathetic nervous system has a role³. Additionally, the control of glucagon secretion may contribute to the suppression of blood glucose by CNS leptin⁷⁶. Determining how each of these systems contributes to the various responses to brain nutrient and hormone action, and the specific brain regions and circuits that control the response, will be important for our understanding of the physiology of metabolism.

Within the brain itself, we have an enormous amount left to learn. In addition to resolving the mechanisms through which the leptin and melanocortin systems contribute to various metabolic conditions, perhaps the most important issue to resolve is the cellular specificity of the system. Clearly, some progress has been made in this regard, especially in the arcuate nucleus, for which some genetic reagents are available for use in analysis; however, we desperately need other methods of genetic analysis to examine defined subpopulations of ventromedial-nucleus and arcuate-nucleus neurons, and to clarify the roles of other hypothalamic and brainstem circuits that have been implicated in metabolic control. ■

- Levin, B. E. Neuronal glucose sensing: still a physiological orphan? *Cell Metab.* **6**, 252–254 (2007).
- Levin, B. E. & Sherwin, R. S. Peripheral glucose homeostasis: does brain insulin matter? *J. Clin. Invest.* **121**, 3392–3395 (2011).
- Pocai, A., Obici, S., Schwartz, G. J. & Rossetti, L. A brain–liver circuit regulates glucose homeostasis. *Cell Metab.* **1**, 53–61 (2005).
- Taborsky, G. J. Jr. Islets have a lot of nerve! Or do they? *Cell Metab.* **14**, 5–6 (2011).
- Wiesli, P. et al. Acute psychological stress affects glucose concentrations in patients with type 1 diabetes following food intake but not in the fasting state. *Diabetes Care* **28**, 1910–1915 (2005).
- McCowan, K. C., Malhotra, A. & Bistrian, B. R. Stress-induced hyperglycemia. *Crit. Care Clin.* **17**, 107–124 (2001).
- Levin, B. E., Routh, V. H., Kang, L., Sanders, N. M. & Dunn-Meynell, A. A. Neuronal glucosensing: what do we know after 50 years? *Diabetes* **53**, 2521–2528 (2004).
- Evans, M. L. & Sherwin, R. S. Blood glucose and the brain in diabetes: between a rock and a hard place? *Curr. Diab. Rep.* **2**, 101–102 (2002).
- Ritter, S., Li, A. J., Wang, Q. & Dinh, T. T. The value of looking backward: the essential role of the hindbrain in counter regulatory responses to glucose deficit. *Endocrinology* **152**, 4019–4032 (2011).
- Schwartz, M. W. & Porte, D. Jr. Diabetes, obesity, and the brain. *Science* **307**, 375–379 (2005).
- Zhang, Y. et al. Positional cloning of the mouse obese gene and its human homologue. *Nature* **372**, 425–432 (1994).
- Rossetti, L. et al. Short term effects of leptin on hepatic gluconeogenesis and *in vivo* insulin action. *J. Biol. Chem.* **272**, 27758–27763 (1997).
- Liu, L. et al. Intracerebroventricular leptin regulates hepatic but not peripheral glucose fluxes. *J. Biol. Chem.* **273**, 31160–31167 (1998).
- Obici, S., Zhang, B. B., Karkanias, G. & Rossetti, L. Hypothalamic insulin signaling is required for inhibition of glucose production. *Nature Med.* **8**, 1376–1382 (2002).
- Gutierrez-Juarez, R., Obici, S. & Rossetti, L. Melanocortin-independent effects of leptin on hepatic glucose fluxes. *J. Biol. Chem.* **279**, 49704–49715 (2004).
- Pocai, A. et al. Central leptin acutely reverses diet-induced hepatic insulin resistance. *Diabetes* **54**, 3182–3189 (2005).
- Ahima, R. S. Central actions of adipocyte hormones. *Trends Endocrinol. Metab.* **16**, 307–313 (2005).
- D'Alessio, D. A., Sandoval, D. A. & Seeley, R. J. New ways in which GLP-1 can regulate glucose homeostasis. *J. Clin. Invest.* **115**, 3406–3408 (2005).
- Briggs, D. I. & Andrews, Z. B. Metabolic status regulates ghrelin function on energy homeostasis. *Neuroendocrinology* **93**, 48–57 (2011).
- Heppner, K. M., Tong, J., Kirchner, H., Nass, R. & Tschöp, M. H. The ghrelin O-acyltransferase-ghrelin system: a novel regulator of glucose metabolism. *Curr. Opin. Endocrinol. Diabetes Obes.* **18**, 50–55 (2011).
- Kleinridders, A., Konner, A. C. & Bruning, J. C. CNS-targets in control of energy and glucose homeostasis. *Curr. Opin. Pharmacol.* **9**, 794–804 (2009).
- Hribal, M. L., Oriente, F. & Accili, D. Mouse models of insulin resistance. *Am. J. Physiol. Endocrinol. Metab.* **282**, E977–E981 (2002).
- Williams, K. W., Scott, M. M. & Elmquist, J. K. Modulation of the central melanocortin system by leptin, insulin, and serotonin: co-ordinated actions in a dispersed neuronal network. *Eur. J. Pharmacol.* **660**, 2–12 (2011).
- Minokoshi, Y. et al. AMP-kinase regulates food intake by responding to hormonal and nutrient signals in the hypothalamus. *Nature* **428**, 569–574 (2004).
- Pocai, A., Muse, E. D. & Rossetti, L. Did a muscle fuel gauge conquer the brain? *Nature Med.* **12**, 50–51 (2006).
- Cota, D. et al. Hypothalamic mTOR signaling regulates food intake. *Science* **312**, 927–930 (2006).
- Diano, S. & Horvath, T. L. Mitochondrial uncoupling protein 2 (UCP2) in glucose and lipid metabolism. *Trends Mol. Med.* **18**, 52–58 (2012).
- Lam, T. K. et al. Hypothalamic sensing of circulating fatty acids is required for glucose homeostasis. *Nature Med.* **11**, 320–327 (2005).
- Vanpatten, S., Karkanias, G. B., Rossetti, L. & Cohen, D. E. Intracerebroventricular leptin regulates hepatic cholesterol metabolism. *Biochem. J.* **379**, 229–233 (2004).
- Nogueiras, R. et al. The central melanocortin system directly controls peripheral lipid metabolism. *J. Clin. Invest.* **117**, 3475–3488 (2007).
- Burdakov, D. & Alexopoulos, H. Metabolic state signalling through central hypocretin/orexin neurons. *J. Cell. Mol. Med.* **9**, 795–803 (2005).
- Myers, M. G. Jr, Munzberg, H., Leininger, G. M. & Leshan, R. L. The geometry of leptin action in the brain: more complicated than a simple ARC. *Cell Metab.* **9**, 117–123 (2009).
- This review highlights the idea that most LRB neurons in the brain are distinct from the canonical pro-opiomelanocortin and AgRP neurons and lie outside the arcuate nucleus.**
- Xu, Y., Elmquist, J. K. & Fukuda, M. Central nervous control of energy and glucose balance: focus on the central melanocortin system. *Ann. NY Acad. Sci.* **1243**, 1–14 (2011).
- Sadagurski, M. et al. IRS2 signaling in LepR-b neurons suppresses Foxo1 to control energy balance independently of leptin action. *Cell Metab.* **15**, 703–712 (2012).
- Taguchi, A., Wartschow, L. M. & White, M. F. Brain IRS2 signaling coordinates life span and nutrient homeostasis. *Science* **317**, 369–372 (2007).
- Patterson, C. M. et al. Leptin action via LepR-b Tyr1077 contributes to the control of energy balance and female reproduction. *Mol. Metab.* <http://dx.doi.org/10.1016/j.molmet.2012.05.001> (25 July 2012).
- Bates, S. H. et al. STAT3 signaling is required for leptin regulation of energy balance but not reproduction. *Nature* **421**, 856–859 (2003).
- Buettner, C. et al. Critical role of STAT3 in leptin's metabolic actions. *Cell Metab.* **4**, 49–60 (2006).
- Robertson, S. et al. Insufficiency of Janus kinase 2-autonomous leptin receptor signals for most physiologic leptin actions. *Diabetes* **59**, 782–790 (2010).
- Jiang, L. et al. Tyrosine-dependent and -independent actions of leptin receptor in control of energy balance and glucose homeostasis. *Proc. Natl Acad. Sci. USA* **105**, 18619–18624 (2008).
- Barnes, M. B., Lawson, M. A. & Beverly, J. L. Rate of fall in blood glucose and recurrent hypoglycemia affect glucose dynamics and noradrenergic activation in the ventromedial hypothalamus. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **301**, R1815–R1820 (2011).

42. Levin, B. E., Magnan, C., Dunn-Meynell, A. & Le Foll, C. Metabolic sensing and the brain: who, what, where, and how? *Endocrinology* **152**, 2552–2557 (2011).
43. Chan, O., Lawson, M., Zhu, W., Beverly, J. L. & Sherwin, R. S. ATP-sensitive K⁺ channels regulate the release of GABA in the ventromedial hypothalamus during hypoglycemia. *Diabetes* **56**, 1120–1126 (2007).
44. Kang, L. *et al.* Glucokinase is a critical regulator of ventromedial hypothalamic neuronal glucosensing. *Diabetes* **55**, 412–420 (2006).
45. Elias, C. F. *et al.* Chemical characterization of leptin-activated neurons in the rat brain. *J. Comp. Neurol.* **423**, 261–281 (2000).
46. Haque, M. S. *et al.* Role of the sympathetic nervous system and insulin in enhancing glucose uptake in peripheral tissues after intrahypothalamic injection of leptin in rats. *Diabetes* **48**, 1706–1712 (1999).
47. Dhillon, H. *et al.* Leptin directly activates SF1 neurons in the VMH, and this action by leptin is required for normal body-weight homeostasis. *Neuron* **49**, 191–203 (2006).
48. Bingham, N. C., Anderson, K. K., Reuter, A. L., Stallings, N. R. & Parker, K. L. Selective loss of leptin receptors in the ventromedial hypothalamic nucleus results in increased adiposity and a metabolic syndrome. *Endocrinology* **149**, 2138–2148 (2008).
49. Klockener, T. *et al.* High-fat feeding promotes obesity via insulin receptor/P13K-dependent inhibition of SF-1 VMH neurons. *Nature Neurosci.* **14**, 911–918 (2011).
50. Kim, K. W. *et al.* FOXO1 in the ventromedial hypothalamus regulates energy balance. *J. Clin. Invest.* **122**, 2578–2589 (2012).
51. Bergen, H. T., Mizuno, T. M., Taylor, J. & Mobbs, C. V. Hyperphagia and weight gain after gold-thioglucose: relation to hypothalamic neuropeptide Y and proopiomelanocortin. *Endocrinology* **139**, 4483–4488 (1998).
52. Coppari, R. *et al.* The hypothalamic arcuate nucleus: a key site for mediating leptin's effects on glucose homeostasis and locomotor activity. *Cell Metab.* **1**, 63–72 (2005).
- This article reports the reactivation of LRB specifically in the arcuate nucleus; although this only slightly altered food intake and adiposity, glucose homeostasis was markedly improved, suggesting an important role for arcuate-nucleus leptin action in the control of glucose homeostasis independently of energy balance.**
53. Morton, G. J. *et al.* Arcuate nucleus-specific leptin receptor gene therapy attenuates the obesity phenotype of Koletsky (*fa^h/fa^h*) rats. *Endocrinology* **144**, 2016–2024 (2003).
54. Morton, G. J. *et al.* Leptin regulates insulin sensitivity via phosphatidylinositol-3-OH kinase signaling in mediobasal hypothalamic neurons. *Cell Metab.* **2**, 411–420 (2005).
- Using molecular–viral and pharmacological approaches, the authors show that insulin-dependent signalling pathways in the arcuate nucleus are crucial for the control of glucose homeostasis by leptin.**
55. Schwartz, M. W. Central nervous system regulation of food intake. *Obesity (Silver Spring)* **14**, 1S–8S (2006).
56. Vong, L. *et al.* Leptin action on GABAergic neurons prevents obesity and reduces inhibitory tone to POMC neurons. *Neuron* **71**, 142–154 (2011).
57. Hill, J. W. *et al.* Direct insulin and leptin action on pro-opiomelanocortin neurons is required for normal glucose homeostasis and fertility. *Cell Metab.* **11**, 286–297 (2010).
58. Lin, H. V. *et al.* Divergent regulation of energy expenditure and hepatic glucose production by insulin receptor in agouti-related protein and POMC neurons. *Diabetes* **59**, 337–346 (2010).
59. Padilla, S. L., Carmody, J. S. & Zeltser, L. M. *Pomc*-expressing progenitors give rise to antagonistic neuronal populations in hypothalamic feeding circuits. *Nature Med.* **16**, 403–405 (2010).
60. Berglund, E. D. *et al.* Direct leptin action on POMC neurons regulates glucose homeostasis and hepatic insulin sensitivity in mice. *J. Clin. Invest.* **122**, 1000–1009 (2012).
61. Huo, L. *et al.* Leptin-dependent control of glucose balance and locomotor activity by POMC neurons. *Cell Metab.* **9**, 537–547 (2009).
62. Hill, J. W. *et al.* Phosphatidylinositol 3-kinase signaling in hypothalamic proopiomelanocortin neurons contributes to the regulation of glucose homeostasis. *Endocrinology* **150**, 4874–4882 (2009).
63. Perez-Tilve, D. *et al.* Melanocortin signaling in the CNS directly regulates circulating cholesterol. *Nature Neurosci.* **13**, 877–882 (2010).
64. Shrestha, Y. B. *et al.* Central melanocortin stimulation increases phosphorylated perilipin A and hormone-sensitive lipase in adipose tissues. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **299**, R140–R149 (2010).
65. Huszar, D. *et al.* Targeted disruption of the melanocortin-4 receptor results in obesity in mice. *Cell* **88**, 131–141 (1997).
66. Ste, M. L., Miura, G. I., Marsh, D. J., Yagaloff, K. & Palmiter, R. D. A metabolic defect promotes obesity in mice lacking melanocortin-4 receptors. *Proc. Natl Acad. Sci. USA* **97**, 12339–12344 (2000).
67. Butler, A. A. *et al.* A unique metabolic syndrome causes obesity in the melanocortin-3 receptor-deficient mouse. *Endocrinology* **141**, 3518–3521 (2000).
68. Sutton, G. M. *et al.* Diet–genotype interactions in the development of the obese, insulin-resistant phenotype of C57BL/6J mice lacking melanocortin-3 or -4 receptors. *Endocrinology* **147**, 2183–2196 (2006).
69. Chen, A. S. *et al.* Inactivation of the mouse melanocortin-3 receptor results in increased fat mass and reduced lean body mass. *Nature Genet.* **26**, 97–102 (2000).
70. Butler, A. A. The melanocortin system and energy balance. *Peptides* **27**, 281–290 (2006).
71. Grill, H. J. Distributed neural control of energy balance: contributions from hindbrain and hypothalamus. *Obesity (Silver Spring)* **14**, 216S–221S (2006).
72. de Backer, M. W. *et al.* Melanocortin receptor-mediated effects on obesity are distributed over specific hypothalamic regions. *Int. J. Obes. (Lond.)* **35**, 629–641 (2011).
73. Cherrington, A. D. The role of hepatic insulin receptors in the regulation of glucose production. *J. Clin. Invest.* **115**, 1136–1139 (2005).
74. Shimomura, I., Hammer, R. E., Ikemoto, S., Brown, M. S. & Goldstein, J. L. Leptin reverses insulin resistance and diabetes mellitus in mice with congenital lipodystrophy. *Nature* **401**, 73–76 (1999).
75. Oral, E. A. *et al.* Leptin-replacement therapy for lipodystrophy. *N. Engl. J. Med.* **346**, 570–578 (2002).
76. Wang, M. Y. *et al.* Leptin therapy in insulin-deficient type I diabetes. *Proc. Natl Acad. Sci. USA* **107**, 4813–4819 (2010).
77. German, J. P. *et al.* Leptin deficiency causes insulin resistance induced by uncontrolled diabetes. *Diabetes* **59**, 1626–1634 (2010).
- In this article, the authors show that brain leptin injection — in the absence of endogenous or exogenous insulin — suffices to normalize blood glucose.**
78. Grill, H. J., Ginsberg, A. B., Seeley, R. J. & Kaplan, J. M. Brainstem application of melanocortin receptor ligands produces long-lasting effects on feeding and body weight. *J. Neurosci.* **18**, 10128–10135 (1998).
79. Williams, D. L., Kaplan, J. M. & Grill, H. J. The role of the dorsal vagal complex and the vagus nerve in feeding effects of melanocortin-3/4 receptor stimulation. *Endocrinology* **141**, 1332–1337 (2000).
80. Boghossian, S., Park, M. & York, D. A. Melanocortin activity in the amygdala controls appetite for dietary fat. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **298**, R385–R393 (2010).
81. Balthasar, N. *et al.* Divergence of melanocortin pathways in the control of food intake and energy expenditure. *Cell* **123**, 493–505 (2005).
82. Rossi, J. *et al.* Melanocortin-4 receptors expressed by cholinergic neurons regulate energy balance and glucose homeostasis. *Cell Metab.* **13**, 195–204 (2011).
83. Begriche, K. *et al.* Melanocortin-3 receptors are involved in adaptation to restricted feeding. *Genes Brain Behav.* **11**, 291–302 (2012).

Acknowledgements The authors thank members of the Myers and Olson laboratories for discussions and scientific insight. M.G.M. is supported by the Marilyn H. Vincent Foundation and by grants from the National Institutes of Health and the American Heart Association.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at go.nature.com/pq166k. Correspondence should be addressed to M.G.M. (mgmyers@umich.edu).

How cancer metabolism is tuned for proliferation and vulnerable to disruption

Almut Schulze¹ & Adrian L. Harris²

Cancer metabolism has received a substantial amount of interest over the past decade. The advances in analytical tools have, along with the rapid progress of cancer genomics, generated an increasingly complex understanding of metabolic reprogramming in cancer. As numerous connections between oncogenic signalling pathways and metabolic activities emerge, the importance of metabolic reprogramming in cancer is being increasingly recognized. The identification of metabolic weaknesses of cancer cells has been used to create strategies for treating cancer, but there are still challenges to be faced in bringing the drugs that target cancer metabolism to the clinic.

By the mid-twentieth century, cancer cells were known to show characteristic alterations in their metabolic activity. These early studies resulted in the hypothesis that irreversible inactivation of respiration is causally involved in the development of tumours¹. Later, increased rates of glutaminolysis and lipid synthesis were found in tumour tissue, and the close association between cancer-cell metabolism and hypoxia was established (reviewed in ref. 2).

Over the past decade, a more complex picture of cancer-cell metabolism has emerged. Many cancers show increased glucose uptake and enhanced glycolytic rates, suggesting that metabolic alteration provides a growth advantage for tumour cells³. Some of these changes are similar to the metabolic response of non-transformed cells to growth-promoting signals, so it is not entirely clear whether these metabolic alterations are specific to cancer or just reflect the increased proliferation of tumour cells. However, different oncogenic signalling pathways target distinct components of the metabolic network. Moreover, tumours with the same genetic lesions have different metabolic profiles depending on the tissue they arise in⁴, suggesting that the tissue environment strongly affects the metabolic activity of cancer cells. Altered metabolic activity is crucial for supporting uncontrolled proliferation, evasion of growth-inhibitory signals, cell migration and the dissemination of metastatic cells into distant tissues. However, metabolic reprogramming also renders cancer cells more susceptible to perturbations within the metabolic network. Identifying these metabolic dependencies could open a window of opportunity for therapeutic intervention.

Oncogenic signalling drives metabolic reprogramming

Cancer cells need to generate large amounts of precursors for macromolecule biosynthesis to allow the accumulation of biomass during cell growth and proliferation (Fig. 1). Enhanced uptake of glucose supports the production of intermediates for the synthesis of lipids, proteins and nucleic acids⁵. In addition, cancer cells have increased glutamine uptake and glutaminolysis, which replenish intermediates of the tricarboxylic acid (TCA) cycle that are redirected into biosynthetic reactions — a process known as anaplerosis⁶.

Oncogenic signalling drives many of the same pathways that are responsible for the metabolic response of normal cells to growth-promoting signals. Activation of AKT by phosphatidylinositol-3-OH kinase (PI(3)K) results in increased glucose uptake, enhanced activity

and mitochondrial localization of hexokinase and increased glycolytic flux. The mammalian target of rapamycin complex 1 (mTORC1) and hypoxia-inducible factor (HIF) (discussed in more detail later) also contribute to the increased expression and activity of glycolytic enzymes². Oncogenic levels of MYC have been linked to increased glutaminolysis through a coordinated transcriptional program that results in glutamine addiction of MYC-transformed cells⁷. MYC also promotes the alternative splicing of the pyruvate kinase gene *PKM*, resulting in enhanced expression of the embryonic isoform PKM2 (ref. 8). PKM2 is highly expressed in rapidly proliferating tissues, and many cancer cells exclusively express this isoform. In contrast to other isoforms, PKM2 can switch from a tetrameric to a dimeric form with lower activity. This switch can be induced in response to tyrosine kinase signalling⁹ and allows the accumulation of glycolytic intermediates for biosynthetic processes.

Tumour suppressor pathways also affect metabolism. For example, *TP53* maintains mitochondrial activity through the expression of cytochrome *c* oxidase 2, and loss of this gene recapitulates the metabolic consequences of the Warburg effect¹⁰. *TP53* regulates glycolysis by inducing the expression of the *TP53*-induced glycolysis and apoptosis regulator (TIGAR), an enzyme with homology to fructose-2,6-bisphosphatase (ref. 11). Increased expression of this regulator inhibits glycolytic activity and increases the availability of glucose-6-phosphate (G6P) for entry into the oxidative arm of the pentose phosphate pathway (PPP) (Fig. 1), thereby supporting the production of riboses and NADPH for nucleotide biosynthesis as part of the DNA-damage response. However, p53 can also reduce the production of NADPH by inhibiting G6P dehydrogenase, the rate-limiting enzyme of this pathway¹². Recent evidence suggests that the metabolic functions of p53 may be essential for its role as a tumour suppressor, whereas other functions — including induction of cell-cycle arrest and apoptosis — are dispensable¹³.

Mitochondrial metabolism

Contrary to previously held views, most cancer cells retain functional mitochondria. Mitochondria are essential for the synthesis of citrate by the TCA cycle for the production of cytoplasmic acetyl-coenzyme A (CoA), a central source of acetyl groups for lipid synthesis and protein acetylation. A large fraction of nuclear-encoded mitochondrial genes are part of the transcriptional program

¹Gene Expression Analysis Laboratory, Cancer Research UK, London Research Institute, 44 Lincoln's Inn Fields, London WC2A 3LY, UK. ²Cancer Research UK Growth Factor Group, The Weatherall Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Headington, Oxford OX3 9DS, UK.

induced by MYC (ref. 2). MYC-transformed cells have been shown to be highly dependent on AMP-activated protein kinase (AMPK)-related kinase 5 (ARK5) (also known as NUA1), which limits the activity of mTORC1 and maintains the high respiratory capacity of these cells¹⁴. However, oxidative mitochondrial metabolism can be impaired in cancer cells as a result of mutations in components of the TCA cycle or electron transport chain. Moreover, tumour hypoxia inhibits the entry of pyruvate into the TCA cycle and prevents the synthesis of citrate through this route. Under these conditions, reductive carboxylation of glutamine-derived α -ketoglutarate by the NADPH-dependent isoforms of isocitrate dehydrogenase, IDH1 and IDH2, is used to generate citrate for lipid synthesis^{15–17} (Fig. 1). IDH2 is predominantly located within the mitochondria, so mitochondrial function contributes to macromolecule biosynthesis in cancer cells. Mitochondria could also be required to restore cytoplasmic pools of NAD^+ through the malate–aspartate shuttle to support the high glycolytic flux of cancer cells. Therefore, mitochondria can no longer be viewed as inactive bystanders but should be recognized as important organelles, which are actively involved in the transformation process by maintaining the biosynthetic capacity of cancer cells.

Bioenergetics and redox balance

The increased biosynthetic activity of cancer cells requires not only enhanced uptake and conversion of nutrients, but also a corresponding increase in the production of NADPH as a reducing agent for anabolic reactions and to maintain cellular redox balance¹⁸. The reduced enzymatic activity of PKM2 may allow the accumulation of glycolytic intermediates and promote the entry of G6P into the oxidative arm of the PPP for the production of NADPH (Fig. 1). Indeed, inhibition of PKM2 by direct modification in response to oxidative stress increases the production of NADPH and reduced glutathione¹⁹. Furthermore, enhanced expression of PFKFB4, an isoform of the bifunctional enzyme phosphofructokinase 2 (PFK2), is essential to balance glycolytic activity and NADPH synthesis for the production of anti-oxidants in prostate cancer cells²⁰. However, different cancers may depend on additional pathways for NADPH production as KRAS expression in a pancreatic cancer model has been found not to increase activity of the oxidative PPP²¹. Cytoplasmic NADPH can also be produced by the oxidative decarboxylation of malate to pyruvate by malic enzyme 1 (ME1) and the conversion of citrate into α -ketoglutarate by IDH1. Although the exact contribution of these enzymes to NADPH production in cancer is not

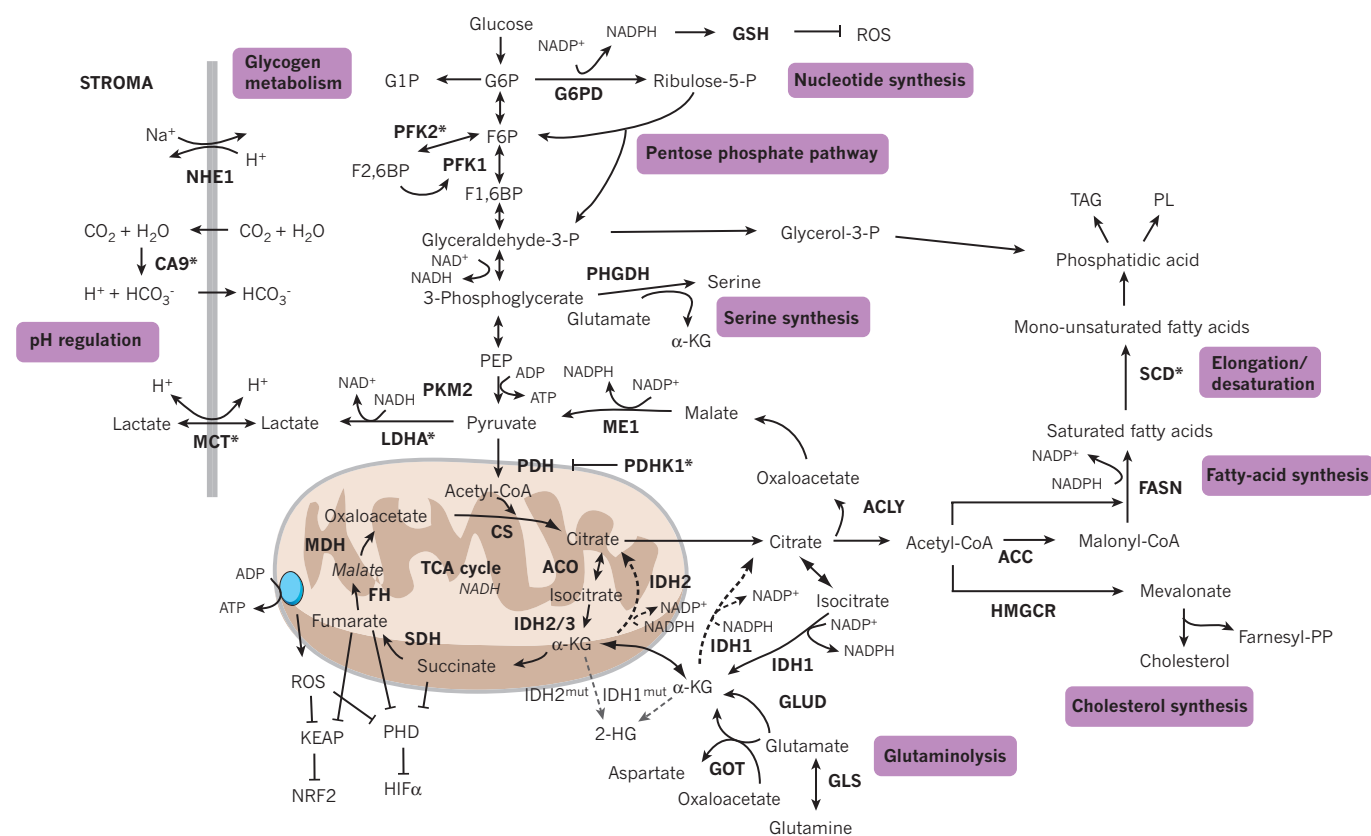


Figure 1 | Overview of metabolic activities in cancer cells. The main metabolic pathways that contribute to the production of macromolecules in mammalian cells are nucleotide synthesis, the pentose phosphate pathway, serine synthesis, glutaminolysis, cholesterol synthesis, fatty-acid synthesis and elongation desaturation. Glycogen synthesis and pH regulation contribute to cellular bioenergetics. The enzymes involved in these pathways are shown in bold, those induced in response to hypoxia are marked with an asterisk. Metabolic enzymes in the TCA cycle, fumarate hydratase (FH) and succinate dehydrogenase (SDH), can act as tumour suppressors. 2-hydroxyglutarate (2-HG) is produced from α -ketoglutarate (α -KG) by the mutant forms of isocitrate dehydrogenase 1 (IDH1) and IDH2 enzymes that are found in cancer (grey dashed arrow). Reductive carboxylation of α -KG by IDH1 and IDH2 produces citrate for lipid synthesis in hypoxic cells (black dashed arrow). ACC, acetyl-CoA carboxylase; ACLY, ATP citrate lyase; ACO, aconitase; CA9, carbonic anhydrase 9; CoA, coenzyme A;

CS, citrate synthase; FASN, fatty-acid synthase; F1,6BP, fructose-1,6-bisphosphate; F2,6BP, fructose-2,6-bisphosphate; F6P, fructose-6-phosphate; GLS, glutaminase; GLUD, glutamate dehydrogenase 1; GOT, glutamic-oxaloacetic transaminase; GSH, glutathione; G1P, glucose-1-phosphate; G6P, glucose-6-phosphate; G6PD, G6P dehydrogenase; HIF, hypoxia inducible factor; HMGCR, 3-hydroxy-3-methylglutaryl-CoA reductase; KEAP, kelch-like ECH-associated protein 1; LDHA, lactate dehydrogenase A; MCT, monocarboxylate transporters; MDH, malate dehydrogenase; ME1, malic enzyme 1; NHE1, Na^+/H^+ exchange protein 1; NRF2, nuclear factor (erythroid-derived 2)-like 2; PDH, pyruvate dehydrogenase; PDHK1, pyruvate dehydrogenase kinase; PEP, phosphoenolpyruvate; PFK, phosphofructokinase; PHD, prolyl hydroxylases; PHGDH, phosphoglycerate dehydrogenase; PKM2, pyruvate kinase M2; PL, phospholipids; ROS, reactive oxygen species; SCD, stearoyl-CoA desaturase; TAG, triacylglycerides.

fully understood, they may present attractive targets for selectively killing tumour cells with high biosynthetic activity.

Lactate transport and pH regulation

Enhanced uptake of glucose and its conversion into lactate creates the problem of intracellular acidification and lactate accumulation. Maintaining an alkaline intracellular milieu is essential for cancer-cell survival, whereas acidification of the extracellular micro-environment may be important to facilitate cancer-cell invasion and metastasis formation (reviewed in ref. 22).

Three main acid-regulatory systems have been implicated in the pH regulation of cancer cells; these involve Na^+/H^+ exchangers, carbonic anhydrase 9 (CA9) and monocarboxylate transporters (Fig. 1). Regulation of intracellular pH by an isoform of an Na^+/H^+ exchanger protein (NHE1) is required for tumour growth, cell migration and metastasis formation²². CA9 is a target gene of HIF and prevents the acidification of cells under hypoxic conditions²³. The catalytic domain of CA9 is located at the extracellular face of the plasma membrane and catalyses the conversion of membrane-permeant carbon dioxide into bicarbonate, thereby removing protons from the cell to maintain an alkaline intracellular milieu.

Lactate transport across the plasma membrane is facilitated by monocarboxylate transporters and is coupled to the symport of protons²⁴. Some monocarboxylate transporters require association with an ancillary protein (CD147, also known as basigin) for membrane localization and activity, and it has been shown that silencing of CD147 in cancer cells prevents transformation and tumour formation because of the inhibition of MCT1 and MCT4 function²⁵. Notably, MCT4 is highly overexpressed in renal cancer and depletion of MCT4 causes accumulation of lactate, acidification and cell death in renal cancer cells²⁶. Renal cancer is associated with loss of the von Hippel-Lindau tumour suppressor protein (pVHL), resulting in stabilization and activation of the α -subunits of HIF (HIF1 α and HIF2 α). The dependency of renal cancer cells on MCT4 is likely to be caused by their high glycolytic activity owing to a pseudohypoxic state. However, the complete picture of lactate transport in cancer may be more complex. Tumour-derived lactate can be taken up and oxidized by stromal cells²⁷, and there is evidence that cells within oxygenated areas of solid tumours may use lactate produced by hypoxic tumour cells²⁸. This use increases glucose availability and supports the survival of cells within the hypoxic part of the tumour. Nevertheless, monocarboxylate transporters are attractive targets for cancer therapy, and inhibitors of MCT1, such as AZD-3965, are currently being tested in clinical trials.

The role of lipid synthesis in cancer

In addition to glucose and glutamine metabolism, the increased biosynthesis of macromolecules — particularly lipids — has been recognized as a component of the metabolic reprogramming in cancer cells²⁹. Although most cells in the adult body rely on lipids from the bloodstream, many cancer cells show a reactivation of *de novo* fatty-acid synthesis³⁰ (Fig. 1). Expression of enzymes involved in the synthesis of cholesterol and fatty acids is controlled by the sterol regulatory element binding proteins (SREBPs). These proteins are activated by AKT in an mTORC1-dependent manner, and SREBP target genes represent one of the main components of the transcriptional program downstream of mTORC1 (refs 31, 32). Inhibition of SREBP function affects cell and organ size in fruitflies (*Drosophila melanogaster*) suggesting that lipid synthesis is essential for cell growth³¹. Furthermore, increased *de novo* fatty-acid synthesis, incorporation of newly synthesized lipids into phosphoglycerides and enhanced expression of SREBP1 have been shown to correlate with breast cancer progression³³.

Although the exact role of lipid synthesis in cancer is not fully understood, it is likely that *de novo* lipogenesis contributes to the generation of structural lipids, such as sterols and phosphoglycerides

that are required for the generation of biological membranes. Triacylglycerides are stored in lipid droplets and can be used to generate energy, whereas some lipids act as second messengers and could contribute to signalling processes in cancer cells. Indeed, monoacylglycerol lipase, an enzyme that is overexpressed in aggressive cancers in humans, induces a specific lipid signature that could trigger signalling events that are involved in cell migration and invasion³⁴.

Enhanced expression of enzymes within the cholesterol biosynthesis (mevalonate) pathway has been shown to induce breast epithelial cells to form three-dimensional structures that may represent early stages of cancer³⁵. p53 was found to associate with SREBPs to drive the expression of these genes³⁵. The cholesterol biosynthesis pathway also provides intermediates for protein isoprenylation and loss of retinoblastoma protein results in enhanced prenylation of NRAS through activation of SREBP³⁶.

The rate-limiting enzyme of this pathway, 3-hydroxy-3-methylglutaryl-CoA synthase (HMGCR), is the molecular target of statins — a group of widely used cholesterol-lowering drugs. HMGCR promotes cancer-cell proliferation and can cooperate with RAS to transform mouse embryo fibroblasts, suggesting that this pathway is crucial for cancer development³⁷. A recent meta-analysis³⁸ failed to establish any beneficial effects of statin use on cancer incidence. However, altered study design, improved pharmacokinetics or the combination of statins with other chemotherapeutic agents may yet demonstrate their potential value for cancer therapy.

Lipid synthesis may also have a non-cell-autonomous role in cancer development. Adipocytes promote growth and metastasis formation of ovarian cancer cells and provide them with lipids for energy generation³⁹. It is also possible that lipogenesis in cancer cells could support the growth of cells located within nutrient-limited areas, thereby contributing to symbiotic relationships within tumours.

Surviving the cancer environment

The micro-environment of many solid tumours is characterized by limited oxygen availability as a result of the distance to the vasculature. HIF drives metabolic adaptation to hypoxic conditions by inducing a distinct transcriptional program⁴⁰. HIF induces the expression of glucose transporters and glycolytic enzymes, including glucose transporter 1 and 3 (GLUT1 and GLUT3), hexokinase 2 and some isoforms of PFK2 (ref. 40). HIF also prevents the entry of pyruvate into the TCA cycle by inducing pyruvate dehydrogenase kinase 1 (PDHK1) (ref. 41, 42) and lowers cellular respiration by regulating cytochrome *c* oxidase isoform expression and inhibiting mitochondrial biogenesis⁴⁰.

Other metabolic constraints imposed by the *in vivo* tumour micro-environment are less well studied. Glucose starvation can select for oncogenic mutations in *KRAS* and thus promote cell transformation⁴³. *In vivo* distribution of other metabolites, including lipids and lipoproteins, could also significantly affect tumour-cell survival and cancer development. Metabolite analysis by magnetic resonance spectroscopy (MRS) and the use of labelled metabolites as tracers for positron emission tomography (PET) are required to establish a more complete picture of the metabolic profile of cancer cells within a solid tumour (Box 1). For example, stable isotope labelling, coupled with nuclear magnetic resonance analysis, has demonstrated that orthotopically implanted human glioblastoma cells use glucose — rather than glutamine — to produce TCA-cycle intermediates⁴⁴. Analysis of cells in the context of intact tissues is clearly required to fully understand the effect of the tumour micro-environment on the metabolic activity of cancer cells.

Oxidative stress and transformation

Metabolic processes have an essential role in the regulation of cellular redox balance. The mitochondrial respiratory chain is the main source of free radicals, mainly through the production of superoxide radicals by complex I and III. Although superoxide produced

BOX 1

Measuring cancer metabolism

Alterations in the metabolic activity of cancer cells can be determined using various techniques. Cancer 'metabolomics' usually describes the comprehensive analysis of the full metabolite composition of cancer cells or tumours, whereas cancer 'metabonomics' is often used to describe the response of a system to perturbation, such as drug treatment. Mass spectrometry allows the detection of a large number of cellular metabolites with high accuracy. It is often combined with initial separation of complex samples by gas or liquid chromatography, or capillary electrophoresis. To identify individual metabolites, mass-spectrometry spectra are compared with data from spectral libraries or reference compounds. However, it is often difficult to identify metabolites that produce overlapping spectral peaks. For accurate metabolite quantification, samples can be spiked with control compounds to determine the efficiency of detection. The technical advances of the past decade have meant the quantitative detection of hundreds or even thousands of metabolites is now routinely possible.

Another technique that has received much interest is the determination of metabolic activity using stable isotope tracing. This technique provides a dynamic representation of the activity of different metabolic processes. Cells are labelled with modified compounds (usually nutrients, such as glucose or glutamine) in which one or several carbon atoms have been replaced with a heavy isotope⁹⁸. Metabolites that incorporate the heavy isotope can be

detected over time by mass spectrometry (through a characteristic shift in their mass) or nuclear magnetic resonance (NMR) spectroscopy.

NMR spectroscopy can also be applied to detect metabolite levels in intact tissues or even in tumours *in vivo*. This generally detects natural abundant heavy isotopes (carbon-13, hydrogen-1, phosphorus-31), but can also be used to follow the metabolism of labelled synthetic metabolic substrates⁴⁴. One limitation is that substrates have to be applied at physiological concentrations, rather than as tracers. However, using hyperpolarized NMR techniques can increase sensitivity⁹⁹.

Another imaging strategy to determine the metabolic activity of tumours is positron emission tomography (PET). This technique detects the accumulation of tracers within the tumour tissue. Fluorodeoxyglucose PET imaging detects primary tumours and metastases owing to their high rate of glucose uptake. Tracers based on other metabolites, such as acetate, glutamine, thymidine or glycine, are in development.

In parallel to analytical advances, mathematical tools for the analysis of metabolism are being developed. Global reconstructions of human metabolic networks are available¹⁰⁰, and metabolic flux analysis can be used to produce quantitative models of metabolic activity. Together, these tools will generate a more comprehensive understanding of the metabolic reprogramming in cancer.

within the mitochondrial matrix can be detoxified through the actions of superoxide dismutase and catalase, free radicals released into the intermembrane space can contribute to the generation of cytoplasmic reactive oxygen species (ROS). Mitochondrial ROS production is required for RAS-dependent cell transformation⁴⁵ and contributes to hypoxic stabilisation of HIF⁴⁶. Inhibition of MYC-dependent expression of mitochondrial genes by FOXO3a modulates ROS metabolism and prevents HIF stabilisation in hypoxic cells⁴⁷. However, expression of physiological levels of oncogenic KRAS (KRAS^{G12D}) in mouse embryo fibroblasts actually lowers ROS levels by activating the nuclear factor (erythroid-derived 2)-like 2 NRF2, which is involved in the regulation of enzymes that detoxify free radicals⁴⁸.

ROS levels can directly affect the metabolic activity of cancer cells. Inhibition of PKM2 by ROS-dependent covalent modification of a cysteine residue supports the production of NADPH for anti-oxidant synthesis by allowing the accumulation of glycolytic intermediates⁴⁹. However, excess ROS production leads to cell death and limits tumour growth. Cancer cells need an optimal level of ROS that supports cancer-promoting signalling functions, but does not lead to irreversible oxidative damage that results in cell death or senescence. Tipping this balance could offer routes into developing cancer therapeutics.

Metabolic oncogenes and tumour suppressors

In addition to the role metabolic alterations have in facilitating the growth-promoting response to oncogene activation, they can also actively drive the transformation process. This revised view of cancer metabolism emerged from the recognition that metabolic enzymes are themselves subject to genetic alterations in cancer.

Mitochondrial tumour suppressors

The discovery that inherited mutations in genes that encode succinate dehydrogenase (*SDHB*, *SDHC* and *SDHD*) or fumarate hydratase (*FH*) are associated with familial cancer syndromes raised

interest in the role of these genes as tumour suppressors (reviewed in ref. 50). The enzymes encoded by these genes are components of the TCA-cycle.

Inhibition of succinate dehydrogenase leads to the inhibition of prolyl hydroxylases by the accumulation of the TCA cycle intermediate succinate. Prolyl hydroxylases are responsible for the hydroxylation of HIF on two proline residues, which labels the protein for VHL-dependent ubiquitylation and subsequent degradation in tissues with a normal oxygen concentration. Prolyl hydroxylases use the oxidative decarboxylation of α -ketoglutarate to transfer a hydroxyl group onto their substrates. This reaction is inhibited in the presence of succinate or fumarate and can result in the accumulation of prolyl hydroxylase substrates, including HIF⁵⁰. Loss of succinate dehydrogenase or fumarate hydratase is associated with certain forms of hereditary renal cancer, and accumulation of succinate as a result of feedback inhibition and stabilization of HIF could lead to cancer development. Recent evidence suggests that renal cyst formation after *Fh1* deletion is independent of HIF, but involves activation of the NRF2 pathway by fumarate, and that activation of NRF2 may contribute to the development of fumarate-hydratase-deficient cancers⁵¹. However, altered epigenetic regulation owing to the inhibition of other α -ketoglutarate-dependent dioxygenases (including histone and DNA demethylases, discussed in more detail later) may also contribute to tumorigenesis in these syndromes⁵².

Oncogenic mutation of metabolic enzymes

Somatic mutations in IDH1 and IDH2 have been found at high frequency in secondary glioblastoma^{53,54} and in acute myeloid leukaemia (AML)⁵⁵. These mutations always cause a single amino-acid change in one of the two alleles of either gene (arginine 132 in IDH1 or arginine 172 in IDH2). Initially, mutation of IDH1 in glioma was thought to lead to dominant-negative inhibition of the wild-type protein and cause activation of HIF through the decreased availability of α -ketoglutarate⁵⁶. However, mutant IDH proteins have been subsequently shown to acquire a neomorphic enzymatic

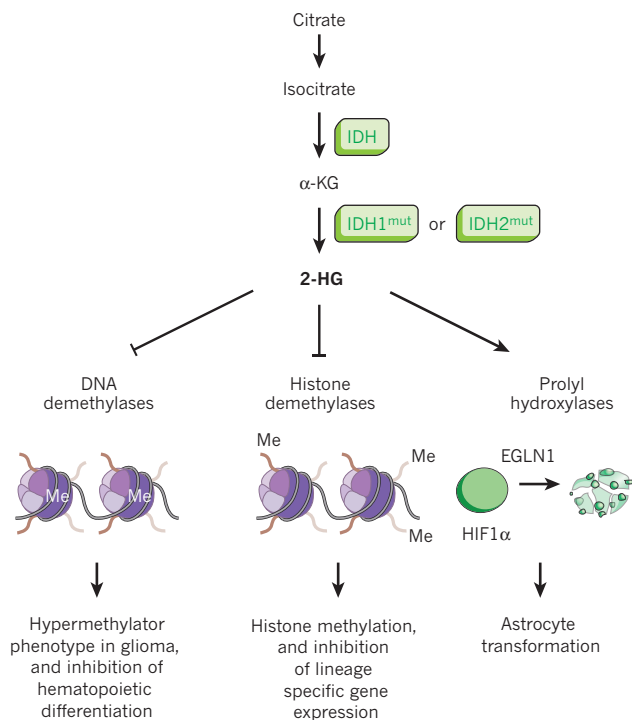


Figure 2 | Epigenetic regulation by 2-hydroxyglutarate. Mutant forms of isocitrate dehydrogenase (IDH) 1 or 2 found in cancer produce the oncometabolite 2-hydroxyglutarate (2-HG) through a neomorphic enzymatic activity. 2-Hydroxyglutarate inhibits DNA demethylases, including TET2, inhibiting haematopoietic differentiation⁶⁰. Inhibition of DNA demethylases by 2-deoxyglucose is also associated with increased genome-wide DNA methylation and a CpG island methylator phenotype in glioma^{63,61}. 2-Hydroxyglutarate inhibits histone demethylases, including lysine-specific demethylase 4C (KDM4C), lysine-specific demethylase 7A (KDM7A) and lysine-specific demethylase 4A (KDM4A). This causes increased histone methylation in glioma cells⁶¹ and leads to the inhibition of expression of lineage-specific differentiation genes in primary astrocytes⁶⁴. The (*R*)-enantiomer of 2-hydroxyglutarate stimulates the activity of the prolylhydroxylase EGLN1, which results in enhanced degradation of HIF1α and promotes transformation in human astrocytes⁶⁶. α-KG, α-ketoglutarate.

activity and catalyse the conversion of α-ketoglutarate to 2-hydroxyglutarate⁵⁷. Indeed, increased levels of 2-hydroxyglutarate have been found in IDH-mutant tumours^{57–59}. This discovery suggested that small molecules could potentially be drivers of cell transformation. The term oncometabolite was coined when it was shown that 2-hydroxyglutarate modulates gene expression in cancer cells by affecting epigenetic regulation (Fig. 2). Mutations of IDH1 and IDH2 were found to be associated with high levels of DNA methylation in AML (ref. 60). Accumulation of 2-hydroxyglutarate inhibits the function of α-ketoglutarate-dependent dioxygenases, including tet methylcytosine dioxygenase 2 (TET2), a DNA demethylase involved in epigenetic regulation⁶¹. Interestingly, mutations of *TET2* in AML are restricted to patients that carry no IDH mutations⁶⁰. 2-Hydroxyglutarate can also affect histone modification by regulating the Jumoni-domain-containing protein 2A (JMJD2A, also known as KDM4A)⁶². IDH1 mutations have also been associated with a hypermethylator phenotype in glioma⁶³, and expression of mutant IDH1 has been shown to lead to the repression of lineage-specific differentiation genes in primary astrocytes⁶⁴ (Fig. 2). These findings suggest that accumulation of 2-hydroxyglutarate as a result of IDH mutation keeps cells in an undifferentiated, or stem-cell-like, state that may be more permissive for transformation. Moreover, a conditional knock-in of mutant IDH1^{R132H} caused an expansion of early haematopoietic progenitors and was associated with methylation changes similar to those observed in human AML⁶⁵.

2-Hydroxyglutarate produced by mutant IDH consists exclusively of the (*R*)-enantiomer (one of two possible stereoisomers). (*R*)-2-Hydroxyglutarate was found to stimulate the activity of the prolyl hydroxylase EGLN1 in human astrocytes (Fig. 2). This resulted in the degradation of HIF and the promotion of cell transformation⁶⁶. Although this result was initially surprising, it was supported by the observation that IDH-mutant gliomas express lower levels of HIF target genes, suggesting that HIF has a tumour suppressive role in this subset of tumours⁶⁶.

Many cellular processes are highly dependent on the availability of metabolites, so metabolic reprogramming may have even more of a marked effect in cancer cells. Post-translational modification of proteins — particularly histones — through acetylation requires acetyl-CoA, and inhibition of acetyl-CoA production can affect histone acetylation in response to growth-factor stimulation⁶⁷. One example of this effect is the sirtuins, a class of NAD⁺-dependent protein deacetylases, which are important regulators of energy metabolism and stress resistance. Changes in the availability of acetyl-CoA or NAD⁺ caused by metabolic reprogramming in cancer may thus affect numerous cellular processes, including gene expression.

The complex role of 2-hydroxyglutarate suggests that altered metabolic activity can have extensive effects on transcriptional and epigenetic regulation in cancer cells. The analysis of complete cancer metabolomes will probably uncover additional oncometabolites with similarly complex functions.

Getting to the clinic

Metabolic reprogramming in cancer is now widely recognized as important in offering opportunities for cancer treatment; however, strategies that target the enhanced glycolytic activity in cancer have, so far, not been very successful as treatments (reviewed in refs 68 and 69). The early antimetabolite 2-deoxyglucose had only limited clinical success as a single-agent treatment, mainly because of a lack of efficacy. One mechanism of resistance to this compound is the induction of autophagy, a mechanism of self-degradation of cellular components. In addition, dose-limiting toxicities that affect brain and cardiac function have been observed, reflecting drug uptake by normal tissue in these key organs⁷⁰.

There is a clear need to explore other metabolic pathways, and potential synthetic lethality, that may be more successful. The increased metabolic activity of cancer cells, which is often associated with hypoxia, should render them selectively sensitive to perturbations within the metabolic network.

Finding new targets

Metabolic reprogramming may render cancer cells highly dependent on specific metabolic enzymes or processes that could be exploited for cancer therapy. However, the search for suitable targets may be complicated by the high plasticity of the metabolic network that can induce compensatory biosynthetic routes to generate the limiting metabolites, as well as the exchange of metabolites between cancer cells and the surrounding tissues. Several studies have used RNA interference screening tools to identify metabolic weaknesses in cancer cells. One study identified phosphoglycerate dehydrogenase, which catalyses the first committed step within the serine biosynthesis pathway, as essential for the *in vivo* growth of breast cancer cells⁷¹. The gene encoding this enzyme, *PHGDH*, lies in a region of frequent copy-number gain in breast cancer and melanomas, suggesting that this enzyme has an important role in the development of these cancers. An independent study found that phosphoglycerate dehydrogenase is responsible for the enhanced diversion of carbon into the serine biosynthesis pathway in cancer cells⁷². Increased serine biosynthesis could provide additional α-ketoglutarate for anaplerosis in cancer cells or be involved in the generation of glycine for nucleotide biosynthesis and cysteine for the production of glutathione. Enhanced flux of glycolytic intermediates into the serine

biosynthesis pathway, owing to the reduced pyruvate kinase activity in PKM2-expressing cancer cells, can also be important to maintain mTORC1 activity under conditions of serine depletion⁷³. A study using a screening approach identified an isoform of phosphofructokinase 2, called PFKFB4, as an important enzyme for the survival of prostate cancer cells²⁰.

An additional challenge is to identify those metabolic processes that specifically support the survival of cancer cells that carry defined oncogenic drivers. For example, AMP-activated kinase family member 5 (ARK5) was found to be essential for the viability of cells that expressed oncogenic levels of MYC (ref. 14). ARK5 was required to limit mTORC1-dependent protein synthesis and maintain mitochondrial activity and glutamine metabolism. Another study used a chemical synthetic-lethal screen to show that inhibition of GLUT1 selectively kills VHL-deficient renal cancer cells⁷⁴.

Metabolic weaknesses in cancer can also be predicted using an *in silico* approach. This uses stoichiometric models of metabolic networks coupled to metabolic flux balance analysis and constraint-based modelling to generate models of cancer metabolism and to predict which metabolic genes are essential for cancer-cell survival⁷⁵. This method was used to identify metabolic pathways that are essential to support the viability of *FH1*-mutant cancer cells (synthetic lethal)⁷⁶. A highly parallel metabolomics approach established cellular consumption and release profiles of more than 200 metabolites across the NCI-60 panel (the National Cancer Institute's collection of 60 human cancer cell lines)⁷⁷. This study reported that glycine consumption correlates with cellular proliferation rate, and suggests that targeting glycine metabolism could selectively compromise nucleotide biosynthesis in rapidly proliferating cancer cells. These approaches have the potential to differentiate between common events that are essential for the metabolic reprogramming of most cancer types and specific metabolic alterations that apply only to cancers from a specific tissue or genetic background. Improved analytical capacity and high-throughput screening will continue to provide insight into the complexity of cancer metabolism.

Patient selection

Selecting the patients who are most likely to benefit from therapies that target cancer metabolism is an additional challenge. One possibility is to stratify patients on the basis of genetic drivers. Metabolic sensitivities in cancer can depend on the activation state of specific oncogenes. The genetic complexity of cancer is, therefore, also likely to be reflected in the cells' specific metabolic requirements. Understanding this complexity is essential for identifying patients who are most likely to benefit from a specific treatment. For example, loss of p53 sensitizes cells to metformin, an inhibitor of mitochondrial ATP production and an activator of AMPK⁷⁸. Metformin enhances the use of fatty acids for energy production and triggers autophagy⁷⁸. Both processes can be used to provide energy and promote cell survival when nutrients are scarce, but they rely on functional p53.

Another oncogenic driver that is frequently activated in human cancer is KRAS. Expression of oncogenic KRAS was found to be essential for tumour maintenance in a genetic model of pancreatic ductal adenocarcinoma²¹. Inhibition of KRAS expression was accompanied by specific metabolic alterations that demonstrated the role of KRAS in glycosylation by the induction of the hexosamine biosynthesis pathway and nucleotide biosynthesis through the non-oxidative arm of the PPP²¹. Treatments that target these pathways may only be effective when this oncogenic driver is present and could be offered to patients who carry this mutation. Other oncogenic lesions with a strong metabolic signal include activation of *PI3K* (also known as *PIK3CA*), *AKT* and *MYC* and loss of *VHL*. Alterations in these genes should be considered when assessing patient responses to targeted treatment.

The metabolic profile of tumours depends not only on the type of genetic lesion but also on the tissue in which the mutation arises⁴.

This is further complicated by the increasingly recognized genetic intratumour heterogeneity of solid tumours⁷⁹. However, because the metabolic state of cancer cells is strongly affected by the tumour micro-environment, biomarkers for tumour hypoxia could be used for patient selection. Several hypoxia gene signatures have been published, which include many of the genes that are involved in glycolytic metabolism. These have been useful in classifying patients who are likely to benefit from radiotherapy in combination with the hypoxic radiosensitizer nimorazole⁸⁰. Similarly, tumours that are already hypoxic may be more sensitive to further deprivation of oxygen through anti-angiogenic therapy. Importantly, *in vivo* assessment of tumour hypoxia by fluoride-18-fluoroazomycin arabinoside (¹⁸F-FAZA) or misonidazole scans could be vital to select those tumours that may be particularly sensitive to these therapies and to monitor treatment response.

Targeting the tumour stroma

The cells of a tumour are strongly influenced by its stromal component. Evidence suggests that there is an exchange of metabolites between cancer and stromal cells to provide nutrients⁸¹. Stromal cells can have an important role in ROS metabolism within the tumour compartment⁸². This may be particularly important in maintaining the replicative potential of cancer stem cells. Although the metabolic requirements of cancer stem cells have not been investigated, early stem-cell development involves a metabolic switch to glycolysis that is reminiscent of the Warburg effect in cancer⁸³. Cancer stem cells require an environment low in ROS and may therefore be highly dependent on specific metabolic activities. Metabolic perturbations that selectively target cancer stem cells may be particularly effective for improving therapeutic response and preventing cancer recurrence⁸⁴. However, suitable molecular markers are required to identify different types of stroma and to predict treatment response.

Cancer diagnostics and dynamic monitoring of therapy

Metabolic reprogramming in cancer has already been exploited for cancer diagnosis and to monitor treatment response (Box 1). Metabolic processes that are highly active in cancer cells can produce specific by-products that can be detected not only in tumour biopsies but also in blood or urine samples⁸⁵. Enhanced glucose uptake forms the basis of tumour imaging by fluorodeoxyglucose-PET (FDG-PET) and can be used as an early indicator of drug efficacy⁸⁶. However, not all tumours can be detected using FDG-PET imaging. The development of molecular tracers based on other metabolites, such as glutamine, acetate, thymidine or glycine, will offer the possibility of profiling the metabolic state of individual tumours and of monitoring the alterations to their metabolic state during treatment. Nuclear magnetic resonance (NMR) spectroscopy offers a non-invasive method for the detection of selected metabolites *in vivo* and can offer insight into the body's metabolic response to targeted therapy⁸⁷. Importantly, changes in metabolic activity may occur quite rapidly in response to therapy, providing the opportunity to revise therapeutic strategies or add agents. *In vivo* imaging of carbon-13-labelled glucose by dynamic MRS has elucidated many aspects of human brain glutamine and glucose metabolism, and could be used to monitor rapid changes in response to therapy⁸⁸. These should be integrated with the use of drugs to target relevant metabolic pathways (for example, glutaminase inhibitors).

Sparing normal tissue and avoiding toxicity

Several compounds that specifically target cancer metabolism are already approved or under clinical trial, and many more are in preclinical development^{68,89}. One of the greatest challenges that is faced during the development of therapeutic strategies to target cancer metabolism is the possibility of toxic effects on non-cancerous tissues. Toxic side effects may be restricted to metabolic tissues, such as liver, but may also have marked effects on whole

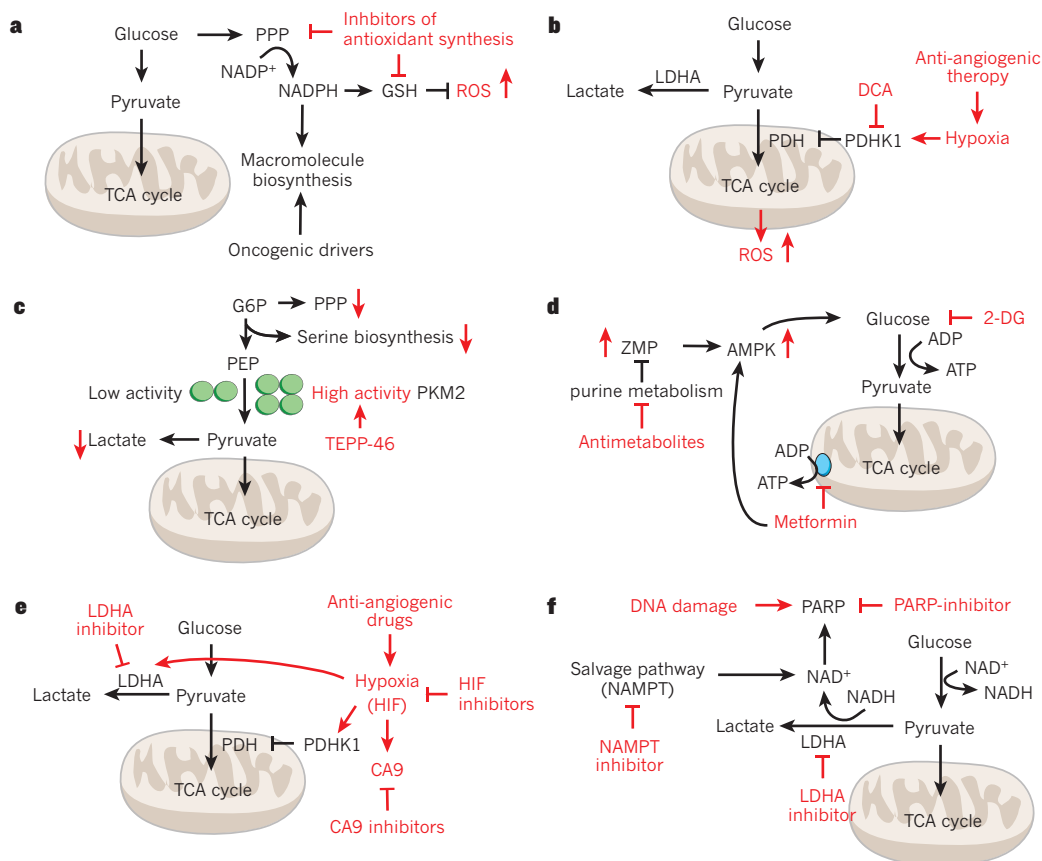


Figure 3 | Exploiting metabolic reprogramming for cancer therapy. **a**, The increased biosynthetic activity of cancer cells renders them highly susceptible to inhibitors of anti-oxidant synthesis. Inhibition of NADPH production by the pentose phosphate pathway (PPP) or disruption of glutathione (GSH) synthesis results in increased reactive oxygen species (ROS). **b**, Reactivation of mitochondrial pyruvate metabolism by inhibition of pyruvate dehydrogenase kinase 1 (PDHK1) with dichloroacetate (DCA) in hypoxic cancer cells or after anti-angiogenic therapy (for example, bevacizumab) can lead to increased production of ROS and cell death. When pyruvate dehydrogenase (PDH) is inhibited, pyruvate is converted into lactate by lactate dehydrogenase A (LDHA). LDHA is also induced by hypoxia. **c**, Activators (such as TEPP-46) of the pyruvate kinase M2 (PKM2) increase glycolytic flux and reduce lactate production in cancer cells. This depletes glycolytic intermediates required for the serine biosynthesis pathway and reduces the availability of glucose-6-phosphate (G6P) for entry into the PPP. **d**, Metformin blocks mitochondrial ATP production, resulting in activation of AMP-activated protein kinase (AMPK) and increased glycolysis. Inhibition of purine

metabolism by antimetabolites such as pemetrexed causes the accumulation of aminoimidazole carboxamide ribonucleotide (ZMP), a cell-intrinsic activator of AMPK. Combination of metformin or pemetrexed with the glycolysis inhibitor 2-deoxyglucose (2-DG) prevents the metabolic adaptation to AMPK activation. **e**, Activation of hypoxia-inducible factor (HIF) in response to anti-angiogenic drugs decreases mitochondrial oxidation of pyruvate through induction of PDHK1, increases the conversion of pyruvate to lactate by inducing expression of LDHA and regulates intracellular pH by inducing carbonic anhydrase 9 (CA9). Combining anti-angiogenic drugs with inhibitors of HIF or inhibitors of its downstream targets (LDHA and CA9) blocks the adaptive response to hypoxia. **f**, Poly-ADP-ribose polymerase (PARP) is activated in response to DNA damaging agents and requires NAD⁺ as a cofactor. Nicotinamide phosphoribosyltransferase (NAMPT) restores cellular NAD⁺ pools through the salvage pathway and is required to maintain glycolytic activity in the presence of LDHA inhibitors. Inhibitors of NAMPT may also synergize with PARP inhibitors, particularly in the treatment of tumours carrying mutations in DNA-damage repair-pathways.

body metabolism. For example, inhibitors of fatty-acid synthase decreased body weight by affecting the hormonal control of food intake in mice⁹⁰.

Whether metabolic reprogramming in cancer is intrinsically different from the metabolic response to proliferative stimuli in non-transformed cells is not clear. Indeed, the proliferative response to receptor ligation in T cells involves increased nutrient uptake and enhanced glycolysis and is mediated by signalling pathways that are frequently activated in cancer, including those of PI(3)K, MYC and HIF (ref. 91). Moreover, some metabolic activities in cancer cells, such as increased glycine metabolism, are strongly correlated with cell proliferation⁷⁷. Inhibition of these processes could also affect other proliferative tissues.

However, because metabolic reprogramming in cancer renders cells highly dependent on ROS metabolism, disrupting the production of antioxidants could provide an effective treatment strategy. Considering the important role of glutathione in ROS detoxification, drugs that block glutathione synthesis, which were

originally developed to overcome drug resistance (for example, L-buthionine-(S,R)-sulfoximine) should be re-evaluated (Fig. 3a). Furthermore, chemotherapeutics that induce oxidative stress could be combined with strategies to block NADPH production to achieve synergistic effects.

Reactivating a suppressed pathway

Particularly relevant is the switch to glycolysis from oxidative phosphorylation by hypoxic induction of PDHK1. PDHK1 is inhibited by dichloroacetate. In patients with glioblastoma, this drug was shown to reactivate mitochondrial function and generate free radicals, which were toxic to tumour growth⁹². A combination of this class of drug with drugs that are activated by hypoxia or those that induce hypoxia could result in synergistic effects (Fig. 3b). Of note was the anti-angiogenic effect after treatment with dichloroacetate⁹². Some of the changes associated with metabolic reprogramming in cancer are also recapitulated in the tumour endothelium⁹³; this tissue may be targeted using similar therapeutic strategies. Another strategy

involving the reactivation of a suppressed pathway would be to reactivate PKM2 in cancer cells using small molecules (Fig. 3c). One of these activators, TEPP-46, decreased the growth of non-small-cell lung cancer cells in a xenograft model⁹⁴.

Blocking the escape route

Metformin is being widely investigated for prevention and treatment of cancer. The drug's main action is through the non-reversible inhibition of complex I of the mitochondrial respiratory chain, leading to a reduction in the ATP:AMP ratio and activation of AMPK. In a preclinical study, a combination of metformin and 2-deoxyglucose was effective in a wide range of tumour types⁹⁵. Antimetabolites that inhibit purine metabolism, such as pemetrexed, result in accumulation of aminoimidazole carboxamide ribonucleotide, an endogenous analogue of AMP and activator of AMPK⁹⁶, and provide the option of combining the inhibition of glycolysis and purine metabolism for synergistic effects (Fig. 3d).

Hypoxia is rapidly induced by anti-angiogenic therapy that targets vascular endothelial growth factor (VEGF) and by vascular targeting agents. Thus, further induction of many of these metabolic pathways will occur, potentially contributing to the survival of tumour cells and their resistance to therapy. Using induced hypoxia to synergize with other drugs, by targeting either HIF itself or the key downstream survival pathways, such as CA9, has proven effective in preclinical models (Fig. 3e). The early assessment of degree of hypoxia induction using PET imaging or CA9 detection could help to classify the responses for personalized intervention.

Synthetic lethality

The most likely way to produce anticancer effects is synthetic lethality, and this follows well-established principles in antibiotic therapy or their combination with chemotherapy. A clear example of synthetic lethal effects is the combination of lactate dehydrogenase A (LDHA) inhibitor with a drug that blocks the synthesis of NAD⁺ through the salvage pathway. NAD⁺ can be synthesised *de novo* or recycled from nicotinamide through the salvage pathway, involving the enzyme nicotinamide phosphoribosyltransferase (NAMPT). Inhibition of LDHA results in the depletion of NAD⁺ — which is crucial to maintain glycolytic flux — and inhibition of NAMPT enhances the effectiveness of LDHA inhibitors⁹⁷ (Fig. 3f). NAMPT inhibitors may also be useful in combination with classical chemotherapeutic agents. Genotoxic damage caused by ionizing radiation or DNA-alkylating drugs leads to the activation of poly-(ADP-ribose) polymerase (PARP) and can rapidly deplete cellular pools of cofactor NAD⁺. Inhibitors of NAMPT block the restoration of NAD⁺ pools and could increase the toxicity of DNA-damaging agents. Inhibition of NAMPT could also be combined with PARP inhibitors, which have been proven to be useful in the treatment of tumours that carry mutations in DNA-repair pathways, such as BRCA-mutant breast cancers. NAMPT inhibitors are currently being tested in phase II clinical trials.

Future research

Elucidating the complex interplay between oncogenic signalling pathways and cellular metabolic activity is an exciting challenge for future research. Metabolic reprogramming of cancer cells can clearly not simply be explained by a shift from oxidative phosphorylation to aerobic glycolysis. However, relatively little is known about the differences in metabolic dependencies of genetically diverse cancer cells or the complex metabolic interactions between tumour and stroma. Although the importance of ROS metabolism in cell transformation and tumour maintenance is becoming more evident, the relative contribution of different metabolic pathways to anti-oxidant production in cancer is not fully understood. Many metabolic pathways involved in the reprogramming of cancer cells are closely linked to the metabolic changes associated with hypoxia. Future research should address how cancer

cells maintain the balance between enhanced biosynthetic activity and the need for antioxidant production. Disrupting this balance should selectively impair the viability of cancer cells and, together with appropriate biomarkers and dynamic cancer imaging, provide new strategies for the treatment of cancer. ■

- Warburg, O. On the origin of cancer cells. *Science* **123**, 309–314 (1956).
- Cairns, R. A., Harris, I. S. & Mak, T. W. Regulation of cancer cell metabolism. *Nature Rev. Cancer* **11**, 85–95 (2011).
- Gatenby, R. A. & Gillies, R. J. Why do cancers have high aerobic glycolysis? *Nature Rev. Cancer* **4**, 891–899 (2004).
- Yuneva, M. O. *et al.* The metabolic profile of tumors depends on both the responsible genetic lesion and tissue type. *Cell Metab.* **15**, 157–170 (2012). **This study found that metabolic alterations, which are associated with tumorigenesis, are dependent on the oncogenic driver and the tissue in which the tumour arises.**
- Vander Heiden, M. G., Cantley, L. C. & Thompson, C. B. Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science* **324**, 1029–1033 (2009).
- DeBerardinis, R. J. *et al.* Beyond aerobic glycolysis: transformed cells can engage in glutamine metabolism that exceeds the requirement for protein and nucleotide synthesis. *Proc. Natl Acad. Sci. USA* **104**, 19345–19350 (2007).
- Wise, D. R. *et al.* Myc regulates a transcriptional program that stimulates mitochondrial glutaminolysis and leads to glutamine addiction. *Proc. Natl Acad. Sci. USA* **105**, 18782–18787 (2008).
- David, C. J., Chen, M., Assanah, M., Canoll, P. & Manley, J. L. HnRNP proteins controlled by c-Myc deregulate pyruvate kinase mRNA splicing in cancer. *Nature* **463**, 364–368 (2010).
- Christofk, H. R., Vander Heiden, M. G., Wu, N., Asara, J. M. & Cantley, L. C. Pyruvate kinase M2 is a phosphotyrosine-binding protein. *Nature* **452**, 181–186 (2008).
- Matoba, S. *et al.* p53 regulates mitochondrial respiration. *Science* **312**, 1650–1653 (2006).
- Bensaad, K. *et al.* TIGAR, a p53-inducible regulator of glycolysis and apoptosis. *Cell* **126**, 107–120 (2006).
- Jiang, P. *et al.* p53 regulates biosynthesis through direct inactivation of glucose-6-phosphate dehydrogenase. *Nature Cell Biol.* **13**, 310–316 (2011).
- Li, T. *et al.* Tumour suppression in the absence of p53-mediated cell-cycle arrest, apoptosis, and senescence. *Cell* **149**, 1269–1283 (2012).
- Liu, L. *et al.* Deregulated MYC expression induces dependence upon AMPK-related kinase 5. *Nature* **483**, 608–612 (2012).
- Wise, D. R. *et al.* Hypoxia promotes isocitrate dehydrogenase-dependent carboxylation of α -ketoglutarate to citrate to support cell growth and viability. *Proc. Natl Acad. Sci. USA* **108**, 19611–19616 (2011).
- Metallo, C. M. *et al.* Reductive glutamine metabolism by IDH1 mediates lipogenesis under hypoxia. *Nature* **481**, 380–384 (2011).
- Mullen, A. R. *et al.* Reductive carboxylation supports growth in tumour cells with defective mitochondria. *Nature* **481**, 385–388 (2011). **References 15 to 17 describe the reductive carboxylation of α -ketoglutarate for the production of citrate.**
- Lunt, S. Y. & Vander Heiden, M. G. Aerobic glycolysis: meeting the metabolic requirements of cell proliferation. *Annu. Rev. Cell Dev. Biol.* **27**, 441–464 (2011).
- Anastasiou, D. *et al.* Inhibition of pyruvate kinase M2 by reactive oxygen species contributes to cellular antioxidant responses. *Science* **334**, 1278–1283 (2011).
- Ros, S. *et al.* Functional metabolic screen identifies 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase 4 (PFKFB4) as an important regulator of prostate cancer cell survival. *Cancer Discov.* **2**, 328–343 (2012).
- Ying, H. *et al.* Oncogenic Kras maintains pancreatic tumors through regulation of anabolic glucose metabolism. *Cell* **149**, 656–670 (2012). **This study used stable isotope labelling and metabolic flux analysis to demonstrate that oncogenic KRAS induces the non-oxidative arm of the pentose phosphate pathway.**
- Chiche, J., Brahimi-Horn, M. C. & Pouyssegur, J. Tumour hypoxia induces a metabolic shift causing acidosis: a common feature in cancer. *J. Cell. Mol. Med.* **14**, 771–794 (2010).
- Swietach, P., Hulikova, A., Vaughan-Jones, R. D. & Harris, A. L. New insights into the physiological role of carbonic anhydrase IX in tumour pH regulation. *Oncogene* **29**, 6509–6521 (2010).
- Halestrap, A. P. & Wilson, M. C. The monocarboxylate transporter family—role and regulation. *IUBMB Life* **64**, 109–119 (2012).
- Le Floch, R. *et al.* CD147 subunit of lactate/H⁺ symporters MCT1 and hypoxia-inducible MCT4 is critical for energetics and growth of glycolytic tumors. *Proc. Natl Acad. Sci. USA* **108**, 16663–16668 (2011).
- Gerlinger, M. *et al.* Genome-wide RNA interference analysis of renal carcinoma survival regulators identifies MCT4 as a Warburg effect metabolic target. *J. Pathol.* **227**, 146–156 (2012).
- Koukourakis, M. I., Giatromanolaki, A., Harris, A. L. & Sivridis, E. Comparison of metabolic pathways between cancer cells and stromal cells in colorectal carcinomas: a metabolic survival role for tumor-associated stroma. *Cancer Res.* **66**, 632–637 (2006).

28. Sonveaux, P. *et al.* Targeting lactate-fueled respiration selectively kills hypoxic tumor cells in mice. *J. Clin. Invest.* **118**, 3930–3942 (2008).
29. Santos, C. R. & Schulze, A. Lipid metabolism in cancer. *FEBS J.* **279**, 2610–2623 (2012).
30. Menendez, J. A. & Lupu, R. Fatty acid synthase and the lipogenic phenotype in cancer pathogenesis. *Nature Rev. Cancer* **7**, 763–777 (2007).
31. Porstmann, T. *et al.* SREBP activity is regulated by mTORC1 and contributes to Akt-dependent cell growth. *Cell Metab.* **8**, 224–236 (2008).
32. Duvel, K. *et al.* Activation of a metabolic gene regulatory network downstream of mTOR complex 1. *Mol. Cell* **39**, 171–183 (2010).
33. Hilvo, M. *et al.* Novel therapeutic opportunities offered by characterization of altered membrane lipid metabolism in breast cancer progression. *Cancer Res.* **71**, 3236–3245 (2011).
34. Nomura, D. K. *et al.* Monoacylglycerol lipase regulates a fatty acid network that promotes cancer pathogenesis. *Cell* **140**, 49–61 (2010).
35. Freed-Pastor, W. A. *et al.* Mutant p53 disrupts mammary tissue architecture via the mevalonate pathway. *Cell* **148**, 244–258 (2012).
- This study reports that tumour-associated mutant forms of p53 can bind to SREBP and induce the expression of enzymes within the mevalonate pathway.**
36. Shamma, A. *et al.* Rb regulates DNA damage response and cellular senescence through E2F-dependent suppression of *N-ras* isoprenylation. *Cancer Cell* **15**, 255–269 (2009).
37. Clendening, J. W. *et al.* Dysregulation of the mevalonate pathway promotes transformation. *Proc. Natl Acad. Sci. USA* **107**, 15051–15056 (2010).
38. Emberson, J. R. *et al.* Lack of effect of lowering LDL cholesterol on cancer: meta-analysis of individual data from 175,000 people in 27 randomised trials of statin therapy. *PLoS One* **7**, e29849 (2012).
39. Nieman, K. M. *et al.* Adipocytes promote ovarian cancer metastasis and provide energy for rapid tumor growth. *Nature Med.* **17**, 1498–1503 (2011).
40. Denko, N. C. Hypoxia, HIF1 and glucose metabolism in the solid tumour. *Nature Rev. Cancer* **8**, 705–713 (2008).
41. Papandreou, I., Cairns, R. A., Fontana, L., Lim, A. L. & Denko, N. C. HIF-1 mediates adaptation to hypoxia by actively downregulating mitochondrial oxygen consumption. *Cell Metab.* **3**, 187–197 (2006).
42. Kim, J. W., Tchernyshyov, I., Semenza, G. L. & Dang, C. V. HIF-1-mediated expression of pyruvate dehydrogenase kinase: a metabolic switch required for cellular adaptation to hypoxia. *Cell Metab.* **3**, 177–185 (2006).
43. Yun, J. *et al.* Glucose deprivation contributes to the development of KRAS pathway mutations in tumor cells. *Science* **325**, 1555–1559 (2009).
44. Marin-Valencia, I. *et al.* Analysis of tumor metabolism reveals mitochondrial glucose oxidation in genetically diverse human glioblastomas in the mouse brain *in vivo*. *Cell Metab.* **15**, 827–837 (2012).
45. Weinberg, F. *et al.* Mitochondrial metabolism and ROS generation are essential for Kras-mediated tumorigenicity. *Proc. Natl Acad. Sci. USA* **107**, 8788–8793 (2010).
46. Hamanaka, R. B. & Chandel, N. S. Mitochondrial reactive oxygen species regulate hypoxic signaling. *Curr. Opin. Cell Biol.* **21**, 894–899 (2009).
47. Ferber, E. C. *et al.* FOXO3a regulates reactive oxygen metabolism by inhibiting mitochondrial gene expression. *Cell Death Differ.* **19**, 968–979 (2012).
48. DeNicola, G. M. *et al.* Oncogene-induced Nrf2 transcription promotes ROS detoxification and tumorigenesis. *Nature* **475**, 106–109 (2011).
- This study shows that the expression of oncogenic alleles of KRAS, BRAF or MYC increases ROS detoxification by activating the NRF2-dependent antioxidant programme.**
49. Anastasiou, D. *et al.* Inhibition of pyruvate kinase M2 by reactive oxygen species contributes to antioxidant responses. *Science* **334**, 1278–1283 (2011).
50. Gottlieb, E. & Tomlinson, I. P. Mitochondrial tumour suppressors: a genetic and biochemical update. *Nature Rev. Cancer* **5**, 857–866 (2005).
51. Adam, J. *et al.* Renal cyst formation in *Fh1*-deficient mice is independent of the Hif/Phd pathway: roles for fumarate in KEAP1 succination and Nrf2 signaling. *Cancer Cell* **20**, 524–537 (2011).
52. Xiao, M. *et al.* Inhibition of α -KG-dependent histone and DNA demethylases by fumarate and succinate that are accumulated in mutations of FH and SDH tumor suppressors. *Genes Dev.* **26**, 1326–1338 (2012).
- References 51 and 52 demonstrate that HIF-independent mechanisms involving the activation of NRF2 or the inhibition of α -ketoglutarate-dependent DNA and histone demethylases contribute to tumorigenesis in fumarate-hydratase- and succinate-dehydrogenase-deficient tumours.**
53. Parsons, D. W. *et al.* An integrated genomic analysis of human glioblastoma multiforme. *Science* **321**, 1807–1812 (2008).
54. Yan, H. *et al.* *IDH1* and *IDH2* mutations in gliomas. *N. Engl. J. Med.* **360**, 765–773 (2009).
55. Mardis, E. R. *et al.* Recurring mutations found by sequencing an acute myeloid leukemia genome. *N. Engl. J. Med.* **361**, 1058–1066 (2009).
56. Zhao, S. *et al.* Glioma-derived mutations in *IDH1* dominantly inhibit *IDH1* catalytic activity and induce HIF-1 α . *Science* **324**, 261–265 (2009).
57. Dang, L. *et al.* Cancer-associated *IDH1* mutations produce 2-hydroxyglutarate. *Nature* **462**, 739–744 (2009).
- This study was the first to describe the neomorphic activity of mutant IDH1 and the production of 2-hydroxyglutarate in cancer.**
58. Ward, P. S. *et al.* The common feature of leukemia-associated *IDH1* and *IDH2* mutations is a neomorphic enzyme activity converting α -ketoglutarate to 2-hydroxyglutarate. *Cancer Cell* **17**, 225–234 (2010).
59. Amary, M. F. *et al.* Ollier disease and Maffucci syndrome are caused by somatic mosaic mutations of *IDH1* and *IDH2*. *Nature Genet.* **43**, 1262–1265 (2011).
60. Figueroa, M. E. *et al.* Leukemic *IDH1* and *IDH2* mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell* **18**, 553–567 (2010).
61. Xu, W. *et al.* Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of α -ketoglutarate-dependent dioxygenases. *Cancer Cell* **19**, 17–30 (2011).
62. Chowdhury, R. *et al.* The oncometabolite 2-hydroxyglutarate inhibits histone lysine demethylases. *EMBO Rep.* **12**, 463–469 (2011).
63. Turcan, S. *et al.* *IDH1* mutation is sufficient to establish the glioma hypermethylator phenotype. *Nature* **483**, 479–483 (2012).
64. Lu, C. *et al.* *IDH* mutation impairs histone demethylation and results in a block to cell differentiation. *Nature* **483**, 474–478 (2012).
65. Sasaki, M. *et al.* *IDH1*(R132H) mutation increases murine haematopoietic progenitors and alters epigenetics. *Nature* **488**, 656–659 (2012).
- References 60 to 65 describe the role of 2-hydroxyglutarate in epigenetic regulation through regulation of α -ketoglutarate-dependent DNA and histone demethylases.**
66. Koivunen, P. *et al.* Transformation by the (R)-enantiomer of 2-hydroxyglutarate linked to EGLN activation. *Nature* **483**, 484–488 (2012).
- This study shows that 2-hydroxyglutarate activates proline hydroxylases and increases the degradation of HIF. This was associated with increased proliferation and transformation of astrocytes.**
67. Wellen, K. E. *et al.* ATP-citrate lyase links cellular metabolism to histone acetylation. *Science* **324**, 1076–1080 (2009).
68. Vander Heiden, M. G. Targeting cancer metabolism: a therapeutic window opens. *Nature Rev. Drug Discov.* **10**, 671–684 (2011).
69. Porporato, P. E., Dhup, S., Dadhich, R. K., Copetti, T. & Sonveaux, P. Anticancer targets in the glycolytic metabolism of tumors: a comprehensive review. *Front. Pharmacol.* **2**, 49 (2011).
70. Stein, M. *et al.* Targeting tumor metabolism with 2-deoxyglucose in patients with castrate-resistant prostate cancer and advanced malignancies. *Prostate* **70**, 1388–1394 (2010).
71. Possemato, R. *et al.* Functional genomics reveal that the serine synthesis pathway is essential in breast cancer. *Nature* **476**, 346–350 (2011).
72. Locasale, J. W. *et al.* Phosphoglycerate dehydrogenase diverts glycolytic flux and contributes to oncogenesis. *Nature Genet.* **43**, 869–874 (2011).
- References 71 and 72 used different strategies to identify the role of serine biosynthesis in supporting cancer-cell growth.**
73. Ye, J. *et al.* Pyruvate kinase M2 promotes *de novo* serine synthesis to sustain mTORC1 activity and cell proliferation. *Proc. Natl Acad. Sci. USA* **109**, 6904–6909 (2012).
74. Chan, D. A. *et al.* Targeting GLUT1 and the Warburg effect in renal cell carcinoma by chemical synthetic lethality. *Sci. Transl. Med.* **3**, 94ra70 (2011).
75. Folger, O. *et al.* Predicting selective drug targets in cancer through metabolic networks. *Mol. Syst. Biol.* **7**, 501 (2011).
76. Frezza, C. *et al.* Haem oxygenase is synthetically lethal with the tumour suppressor fumarate hydratase. *Nature* **477**, 225–228 (2011).
- References 75 and 76 describe the application of metabolic models to predict drug targets in cancer and to identify synthetic-lethal metabolic processes in fumarate-hydratase-deficient tumours.**
77. Jain, M. *et al.* Metabolite profiling identifies a key role for glycine in rapid cancer cell proliferation. *Science* **336**, 1040–1044 (2012).
78. Buzzai, M. *et al.* Systemic treatment with the antidiabetic drug metformin selectively impairs p53-deficient tumor cell growth. *Cancer Res.* **67**, 6745–6752 (2007).
79. Gerlinger, M. *et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N. Engl. J. Med.* **366**, 883–892 (2012).
80. Toustrup, K. *et al.* Gene expression classifier predicts for hypoxic modification of radiotherapy with nimorazole in squamous cell carcinomas of the head and neck. *Radiother. Oncol.* **102**, 122–129 (2012).
81. Whitaker-Menezes, D. *et al.* Evidence for a stromal-epithelial “lactate shuttle” in human tumors: MCT4 is a marker of oxidative stress in cancer-associated fibroblasts. *Cell Cycle* **10**, 1772–1783 (2011).
82. Zhang, W. *et al.* Stromal control of cystine metabolism promotes cancer cell survival in chronic lymphocytic leukaemia. *Nature Cell Biol.* **14**, 276–286 (2012).
83. Zhou, W. *et al.* HIF1 α induced switch from bivalent to exclusively glycolytic metabolism during ESC-to-EpiSC/hESC transition. *EMBO J.* **31**, 2103–2116 (2012).
84. Hirsch, H. A., Iliopoulos, D., Tschlis, P. N. & Struhl, K. Metformin selectively targets cancer stem cells, and acts together with chemotherapy to block tumor growth and prolong remission. *Cancer Res.* **69**, 7507–7511 (2009).
85. Sreekumar, A. *et al.* Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. *Nature* **457**, 910–914 (2009).
86. Chen, Z. *et al.* A murine lung cancer co-clinical trial identifies genetic modifiers of therapeutic response. *Nature* **483**, 613–617 (2012).
87. Belouche-Babari, M. *et al.* Histone deacetylase inhibition increases levels of choline kinase α and phosphocholine facilitating noninvasive imaging in human cancers. *Cancer Res.* **72**, 990–1000 (2012).
88. Rothman, D. L., De Feyter, H. M., de Graaf, R. A., Mason, G. F. & Behar, K. L. 13C MRS studies of neuroenergetics and neurotransmitter cycling in humans. *NMR Biomed.* **24**, 943–957 (2011).
89. Tennant, D. A., Duran, R. V. & Gottlieb, E. Targeting metabolic transformation for cancer therapy. *Nature Rev. Cancer* **10**, 267–277 (2010).

90. Mera, P. *et al.* C75 is converted to C75-CoA in the hypothalamus, where it inhibits carnitine palmitoyltransferase 1 and decreases food intake and body weight. *Biochem. Pharmacol.* **77**, 1084–1095 (2009).
91. Altman, B. J. & Dang, C. V. Normal and cancer cell metabolism: lymphocytes and lymphoma. *FEBS J.* **279**, 2598–2609 (2012).
92. Michelakis, E. D. *et al.* Metabolic modulation of glioblastoma with dichloroacetate. *Sci. Transl. Med.* **2**, 31ra34 (2010).
93. Jain, R. K. & Carmeliet, P. Tumor angiogenesis. *Cell* **149**, 1408 (2012).
94. Anastasiou, D. *et al.* Pyruvate kinase M2 activators promote tetramer formation and suppress tumorigenesis. *Nature Chem. Biol.* **8**, 839–847 (2012).
95. Cheong, J. H. *et al.* Dual inhibition of tumor energy pathway by 2-deoxyglucose and metformin is effective against a broad spectrum of preclinical cancer models. *Mol. Cancer Ther.* **10**, 2350–2362 (2011).
96. Rothbart, S. B., Racanelli, A. C. & Moran, R. G. Pemetrexed indirectly activates the metabolic kinase AMPK in human carcinomas. *Cancer Res.* **70**, 10299–10309 (2010).
97. Le, A. *et al.* Inhibition of lactate dehydrogenase A induces oxidative stress and inhibits tumor progression. *Proc. Natl Acad. Sci. USA* **107**, 2037–2042 (2010).
98. Metallo, C. M., Walther, J. L. & Stephanopoulos, G. Evaluation of ¹³C isotopic tracers for metabolic flux analysis in mammalian cells. *J. Biotechnol.* **144**, 167–174 (2009).
99. Kurhanewicz, J., Bok, R., Nelson, S. J. & Vigneron, D. B. Current and potential applications of clinical ¹³C MR spectroscopy. *J. Nucl. Med.* **49**, 341–344 (2008).
100. Duarte, N. C. *et al.* Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc. Natl Acad. Sci. USA* **104**, 1777–1782 (2007).

Acknowledgements The authors thank C. Santos and S. Ros for their critical reading and feedback. A.S. is funded by Cancer Research UK and the EMBO Young Investigator Programme. A.L.H. is funded by Cancer Research UK, the Oxford Cancer Imaging Centre, the Breast Cancer Research Foundation and the Oxford NIHR Biomedical Research Centre.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at go.nature.com/zgayuj. Correspondence should be addressed to A.S. (almut.schulze@cancer.org.uk).

Mitochondrial disorders as windows into an ancient organelle

Scott B. Vafai^{1,2} & Vamsi K. Mootha^{1,2,3}

Much of our current knowledge about mitochondria has come from studying patients who have respiratory chain disorders. These disorders comprise a large collection of individually rare syndromes, each presenting in a unique and often devastating way. In recent years, there has been great progress in defining their genetic basis, but we still know little about the cascade of events that gives rise to such diverse pathology. Here, we review these disorders and explore them in the context of a contemporary understanding of mitochondrial evolution, biochemistry and genetics. Fully deciphering their pathogenesis is a challenging next step that will inspire the development of drug treatments for rare and common diseases.

The field of mitochondrial medicine began in 1959 when Swedish endocrinologist Rolf Luft and his colleagues described the case of a young woman with euthyroid hypermetabolism, which was characterized by profuse sweating and weight loss despite high calorie intake¹. Muscle biopsy and enzyme analysis — which are now a linchpin for the diagnosis of these disorders — revealed an uncoupling of mitochondria in the patient. Although described decades ago, the condition has been reported only once more² and the root cause of Luft's disease remains a mystery.

Since the initial case report, more than 150 distinct genetic mitochondrial syndromes have been defined. The largest subset arises from lesions that influence the function of the respiratory chain, which affect at least 1 in 5,000 live births³ and are the focus of this Review. These diseases can present either in infancy or adulthood, and in a multisystemic or highly tissue-specific manner. Signature traits can include lactic acidosis, skeletal myopathy, deafness, blindness, subacute neurodegeneration, intestinal dysmotility and peripheral neuropathy. Most organ systems can be affected or spared, in varying combinations (Fig. 1).

The clinical features of mitochondrial disorders have been reviewed in detail^{4,5}, but select examples are illustrative. Leber's hereditary optic neuropathy (LHON) — which is the result of point mutations in a ubiquitously expressed mitochondrial DNA (mtDNA)-encoded respiratory chain protein — demonstrates remarkable tissue specificity, with patients developing sudden vision loss as young adults. By contrast, point mutations in an mtDNA-encoded transfer RNA cause mitochondrial myopathy, encephalopathy, lactic acidosis and stroke-like episodes (MELAS), with a multisystem presentation that includes seizures, impaired hearing, stroke-like episodes and lactic acidosis. Mutations in the nuclear-encoded mitochondrial DNA polymerase gene *POLG* underscore pleiotropy: some mutations result in mild ocular-muscle weakness, whereas others produce Alpers' syndrome (characterized by psychomotor regression, seizures and liver failure). These genetic disorders can be mimicked by drugs that have off-target toxicity. A 1995 phase 2 clinical trial for fialuridine — a promising new antiviral for hepatitis B — was halted after the drug caused lactic acidosis, myopathy, neuropathy and even fatal liver failure in some individuals⁶.

As our ability to define the genetic and environmental bases of mitochondrial disorders has accelerated, new questions have arisen. Why are some disorders highly tissue-specific, and others multisystemic? How can loss-of-function mutations in bacteria-conserved proteins even be compatible with life? Why do some drugs that inhibit mitochondria cause disease, whereas other inhibitors are protective against it? What underlies sporadic cases of mitochondrial disease? The answers to most of these questions are a mystery, but if solved they could provide fundamental insight into metabolism, lead to new therapies for orphan diseases, and help us to rigorously evaluate the role of mitochondria in common diseases.

In this Review, we outline four of the key properties of mitochondria — their evolutionary origins, metabolic interconnections, heterogeneity and robustness — that help to explain some of the mysterious features of these disorders. We review recent insight into their genetic architecture, and emerging themes in their pathogenesis. Finally, we discuss the challenges and opportunities that lie ahead for this relatively young field of medicine.

Origin and evolution of mitochondria

Mitochondria have an endosymbiotic origin and retain many vestiges of their bacterial ancestry, including a double membrane and a circular genome (the mtDNA). They resemble microbes in that they are typically about one micrometre in scale and constantly move, divide and fuse to form a dynamic network. Although mitochondria are referred to as semi-autonomous organelles, billions of years of expansive and reductive evolution (Fig. 2a) — accompanied by transfer of most of their genes to the nuclear genome — have now effectively hard-wired these organelles within eukaryotic cells.

Human mtDNA is maternally inherited and encodes only 13 proteins, as well as the 22 tRNA and 2 ribosomal RNA genes required for their translation. All other proteins required to maintain and express mtDNA are encoded by the nuclear genome. Great progress has been made in defining the mammalian mitochondrial proteome, with more than 1,100 proteins assigned to this compartment⁷. Interestingly, mtDNA has a monophyletic origin⁸, whereas the history of the mitochondrial proteome is far more complex^{7,9}. Of the 1,100 known mitochondrial proteins, about two-thirds have bacterial origins (probably from multiple phyla), with the rest representing

¹Departments of Molecular Biology and Medicine, Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. ²Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA. ³Department of Systems Biology, Harvard Medical School, Boston, Massachusetts 02115, USA.

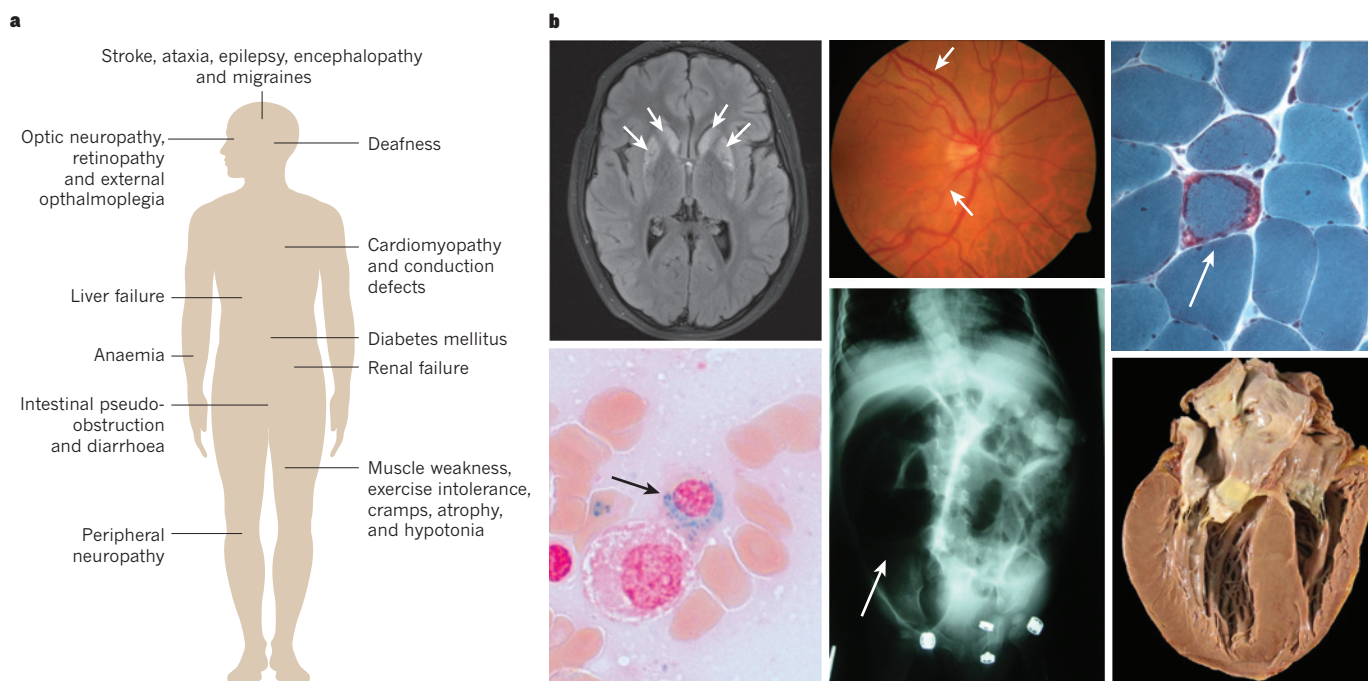


Figure 1 | Phenotypic spectrum of mitochondrial disorders. **a**, Common clinical manifestations of mitochondrial disorders. **b**, Clinical images depicting pathology from patients with a variety of mitochondrial disorders. Clockwise from top left, 3-Tesla fluid-attenuated inversion-recovery brain magnetic resonance imaging demonstrating Leigh syndrome lesions, which are characterized in this image by a hyperintense signal within the caudate and putamen bilaterally (arrows) seen on an axial cut through the basal ganglia; retinal image of the acute phase of Leber's hereditary optic neuropathy, demonstrating an optic disc with swollen nerve fibre layer that is associated with engorged and obscured blood vessels (arrows); ragged red fibre (arrow)

seen on a modified Gomori-trichrome-stained skeletal-muscle section; anterior four-chamber cross-section of a heart that shows signs of hypertrophic cardiomyopathy, including cardiomegaly and asymmetrical septal hypertrophy; plain abdominal radiograph, showing massive bowel distention (arrow) in the setting of chronic intestinal pseudo-obstruction without evidence of mechanical obstruction; and bone-marrow aspirate sample that has been stained for iron to demonstrate a ringed sideroblast (arrow) — characterized by a halo of iron-laden mitochondria around the nucleus of an erythrocyte precursor — from a patient with myopathy, lactic acidosis and sideroblastic anaemia syndrome.

eukaryotic innovations⁷. The mosaic composition of human mitochondria is evident in the organelle's replication and translation machinery, with the ribosome closely resembling its bacterial counterpart¹⁰ and the DNA polymerase resembling that of a viral (bacteriophage) ancestor¹¹. As discussed later, the molecular basis of certain mitochondrial pathologies becomes clear when the ancestry of the organelle is taken into account.

During the course of evolution, the organelle ceded ownership of certain pathways to the rest of the cell, but retained and even acquired others. For example, ribonucleotide reductase — which is used for *de novo* synthesis of deoxyribonucleotides — is found only in the cytosol, and deficiency of this enzyme causes mtDNA depletion syndrome¹². In other cases, the mitochondria have retained a duplicated copy of the cytosolic pathway, such as for tetrahydrofolate-dependent one-carbon metabolism¹³. These paralogous one-carbon pathways seem to have adopted a different functional importance, and may be particularly relevant in disease states such as cancer¹⁴. Understanding the logic of compartmentalization and paralogous pathways is an ongoing challenge.

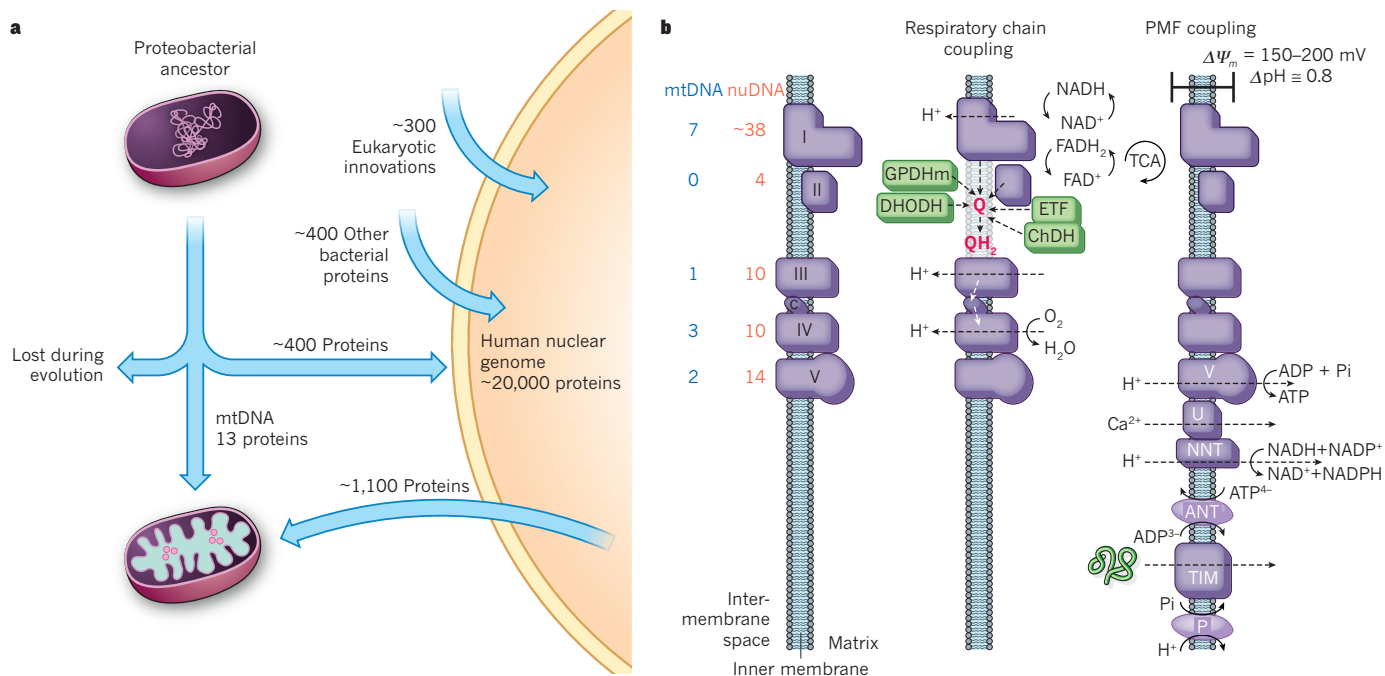
Respiratory chain and its connections

At the heart of mitochondria is the respiratory chain, the core machinery for oxidative phosphorylation (Fig. 2b). Classically, the respiratory chain is defined as four macromolecular complexes that catalyse electron transfer from reducing equivalents, which are derived from intermediary metabolism, to molecular oxygen. Free energy is conserved by coupling electron transport to the formation of a proton gradient, or proton motive force (PMF), by three of these complexes (I, III and IV), which is then dissipated by F_1F_0 -ATPase (complex V) for ATP synthesis. These complexes are associated with the inner membrane and consist of about 90 protein components,

only 13 of which are mtDNA-encoded. Although typically depicted as a linear chain operating in isolation, the respiratory chain is truly a hub in the network of cellular metabolism that is characterized by convergence and divergence of pathways, supercomplex formation and reversibility.

Almost all of the cell's redox reactions ultimately feed into the respiratory chain (Fig. 2b). Complexes I and II mediate two-electron transfer from NADH and $FADH_2$, respectively, to the mobile electron carrier coenzyme Q, providing links to the tricarboxylic acid (TCA) cycle. Coenzyme Q can also receive electrons from *de novo* pyrimidine biosynthesis, fatty-acid and amino-acid oxidation, choline oxidation (ultimately affecting one-carbon metabolism), and glycolysis. Complex III, through its 'Q-cycle', is an adaptor that receives two electrons from reduced coenzyme Q and funnels individual electrons to cytochrome *c*. Complex IV ends the respiratory chain by accepting electrons from cytochrome *c* and using them to fully reduce oxygen to water. Reactive oxygen species (ROS) are potentially toxic by-products of these reactions — especially at complexes I and III — but are buffered by dedicated superoxide dismutase and catalase, as well as glutathione, thioredoxin and protein thiol systems. Interruptions to the respiratory chain can therefore affect nucleotide pools, TCA-cycle flux, one-carbon metabolism and ROS signalling to unleash numerous ripples (discussed later).

The PMF is best known for driving ATP synthesis through oxidative phosphorylation, but it is linked to many other processes (Fig. 2b). The nicotinamide nucleotide transhydrogenase relies on the PMF to regenerate mitochondrial NADPH, which is required for ROS homeostasis. Furthermore, the PMF is coupled to solute and ion transport across the inner membrane, and collapse of the PMF can halt essential biosynthetic reactions, such as Fe-S cluster



the ability to tune their local energetic and redox state through local control of gene expression, the so-called co-location for redox regulation hypothesis²⁷.

Recent proteomic surveys of mitochondria from different organs have quantified the level of molecular heterogeneity, revealing that mitochondria from two distinct organs will typically share about 75% of their components⁷. This tissue heterogeneity has a wide range of functional consequences^{7,28}. For example, whereas most respiratory chain complexes are invariant across organs, complex IV exhibits tissue-specific isoforms⁷ that were previously posited as crucial for responding to local oxygen tension²⁹. The mitochondrial ribosome is used to translate the 13 mtDNA-encoded respiratory chain subunits, all of which are essential and found in all tissues. However, the protein composition of the mitochondrial ribosome exhibits striking tissue diversity⁷, the consequences of which are currently unknown. It is tempting to speculate that this variability contributes to the tissue-specific effect of certain mtDNA mutations.

An exciting new frontier in mitochondrial biology is defining the regulatory mechanisms that give rise to the observed heterogeneity, both within cells and across tissues (Fig. 3b). These mechanisms can be grouped into four broad categories, each of which has been reviewed elsewhere: organelle biogenesis^{30,31}, movement³², fusion and fission³³, and mitophagy³⁴. Co-location for redox regulation may in principle contribute to intracellular heterogeneity, although this hypothesis is in need of rigorous testing, and an understanding of the molecular mechanisms is still lacking. Mutations that affect some of these regulatory programs have already been identified as contributors to disease. For example, mutations in *MFN2* and *OPA1*, both of which are required for mitochondrial fusion, cause Charcot-Marie-Tooth disease type 2A and autosomal dominant optic atrophy, respectively. *PINK1* and *parkin* are required for mitophagy, and are mutated in Mendelian forms of Parkinson's disease.

Robustness

The mitochondrial respiratory chain is a robust system, which is capable of responding to fluctuating nutrient availability and demands. Cardiac mitochondria provide a striking example of robustness: they are capable of maintaining a constant ATP to ADP ratio over a fivefold dynamic range in workload *in vivo* during exercise³⁵. Two mechanisms underlying this regulation have been identified. First, Chance and Williams³⁶, in their pioneering studies of isolated mitochondria, demonstrated that under many experimental conditions the rate of respiration and ATP synthesis is largely controlled by the availability of ADP — a feedback mechanism they termed respiratory control. The second mechanism posits that calcium operates in a feed-forward manner to ensure matched ATP use in the cytosol and its production in mitochondria³⁷. Many cytosolic processes, such as neurotransmission and muscle contraction, are triggered by a rise in cytosolic calcium, and the same calcium signal can be transmitted into the matrix through the uniporter to stimulate the TCA cycle to ensure ATP production.

Robustness probably extends across the entire respiratory chain and its many coupled reactions, as evidenced by classic studies of metabolic control analysis (MCA)^{38,39}. MCA provides an experimental and theoretical framework with which to understand the distribution of flux control of a system property. For a system flux J , the control coefficient C'_{E_i} exerted by enzyme E_i is defined as $(\Delta J/J)/(\Delta E_i/E_i)$, in which $\sum C'_{E_i} = 1$. C'_{E_i} denotes the percentage of control exerted by a single enzyme on the system. Experimentally, small molecules can be used to modulate E_i while J is measured. MCA has been used⁴⁰ to show that control of respiration is broadly distributed and, moreover, that the distribution of control depends on workload. An important implication of such work is that the activity of a respiratory chain complex can be varied over a wide regime before overall respiration is affected (Fig. 3c).

Robustness of mitochondrial metabolism has important implications for understanding disease⁴¹. First, the robustness may be the reason why

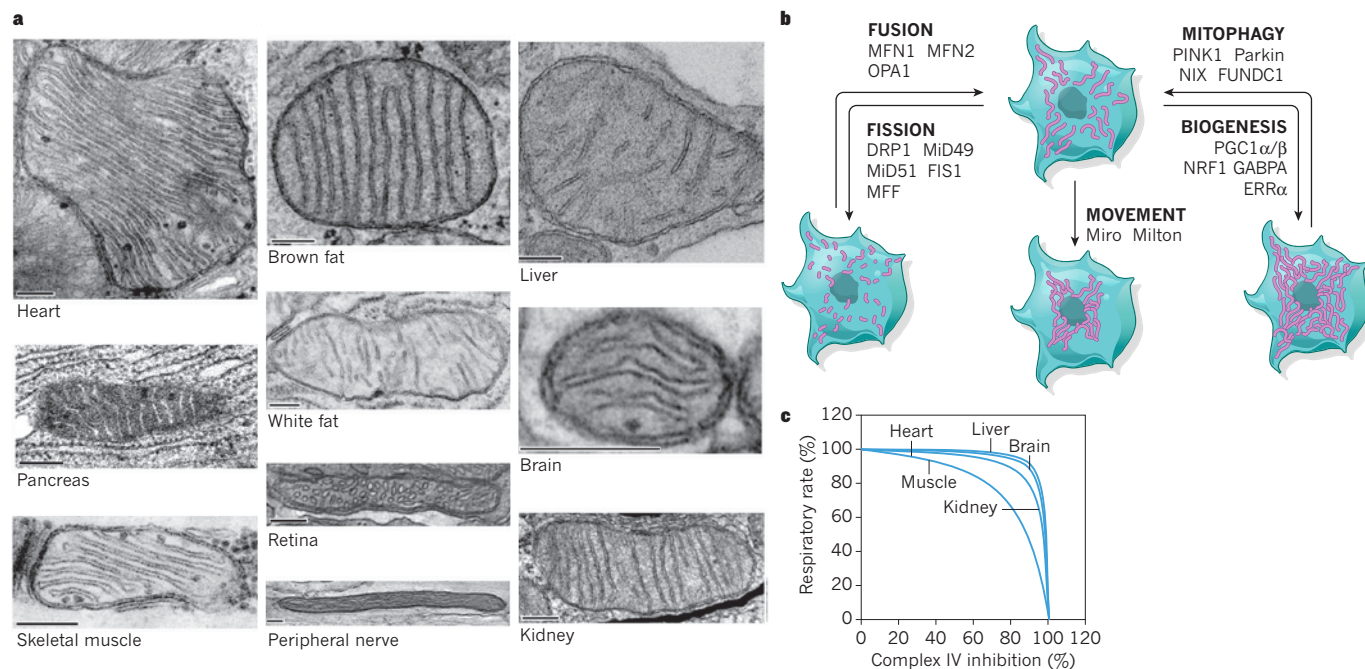


Figure 3 | Mitochondrial tissue heterogeneity and robustness. **a**, Electron micrographs of mitochondria from various tissues, highlighting the diversity across tissues (scale bar, 200 nm). **b**, Processes that give rise to mitochondrial heterogeneity across tissues and within cells: movement, biogenesis, mitophagy, fusion–fission. Key regulators of these processes are listed. DRP1 encoded by *DNM1L*; ERRα encoded by *ESRRA*; FIS1 encoded by *FIS1*; FUNDC1 encoded by *FUNDC1*; GABPA encoded by *GABPA*; MFF encoded by *MFF*; MFN1 encoded by *MFN1*; MFN2 encoded by *MFN2*; MiD49 encoded by *SMCR7*; MiD51 encoded by *SMCR7L*; Milton, which represents the Milton

family of proteins encoded by *TRAK1* and *TRAK2* in humans; Miro, which represents the MIRO family of proteins, encoded by *RHOT1* and *RHOT2* in humans; NIX encoded by *BNIP3L*; NRF1 encoded by *NRF1*; OPA1 encoded by *OPA1*; Parkin encoded by *PARK2*; PGC1α encoded by *PPARGC1A*; PGC1β encoded by *PPARGC1B*; and PINK1 encoded by *PINK1*. **c**, Respiratory rate in isolated mitochondria from five types of rat tissue with complex IV inhibited to varying degrees with KCN. The rate is expressed as a percentage of the respiratory rate observed in untreated mitochondria from that tissue. Panel c adapted with permission from ref. 42.

even severe mutations can be tolerated within the oxidative phosphorylation system and be compatible with life. Robustness of mitochondria implies that the system will be tolerant of hypomorphic alleles, which may be genetic contributors to disease when compounded with the appropriate stress, environmental modifier or a second genetic hit. The control coefficients of different respiratory chain complexes have been shown⁴² to vary across rat tissues (Fig. 3c), leading to speculation that tissue-specific control coefficients have an influence when pathology becomes manifest. However, this attractive hypothesis still needs formal proof.

Genetics of mitochondrial disorders

Over the past 25 years, studies of individual patients and families who are affected by mitochondrial disorders have yielded a wealth of insight into their genetic architecture. These genetic lesions can lie in the mtDNA or nuclear DNA and have revealed the pathways that support respiratory chain assembly and activity (Fig. 4 and Table 1). These genetic studies have also helped to expand the phenotypic spectrum of mitochondrial disorders.

Mutations in the mtDNA

The sequencing of human mtDNA in 1981 (ref. 43) launched the molecular era of mitochondrial medicine. Two landmark papers in 1988 (refs 44 and 45) reported point mutations and deletions in mtDNA in LHON and mitochondrial myopathy, respectively. Since then, more than 300 point mutations, deletions and duplications have been associated with a wide variety of symptoms (<http://www.mitomap.org>). mtDNA is oocyte-derived, so inherited forms of these disorders follow maternal inheritance. Although

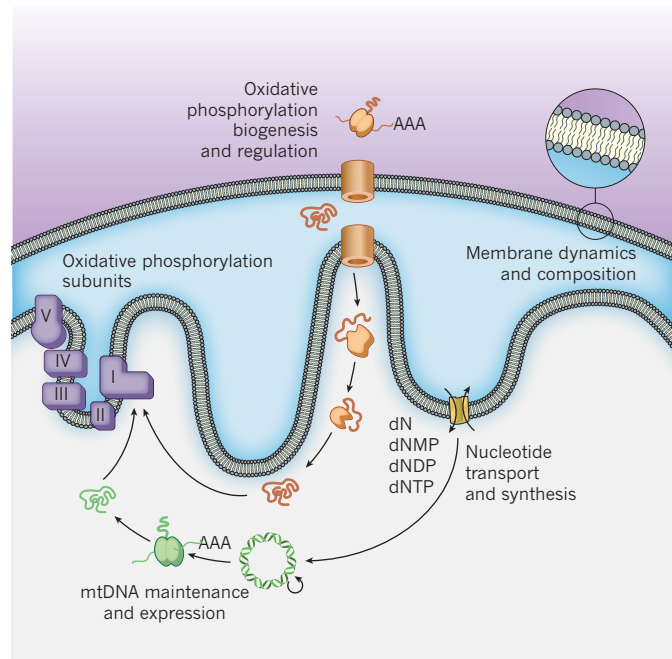


Figure 4 | Genetic pathways underlying mitochondrial respiratory chain disorders. The genes that are known to be mutated in respiratory chain disorders can be grouped in five broad categories on the basis of the pathway in which they participate. The products of genes identified so far are listed in Table 1, and include those involved in disorders of individual oxidative phosphorylation subunits of complexes I–V; proteins involved in mtDNA replication, transcription or translation; proteins involved in the assembly of oxidative phosphorylation complexes, mitochondrial protein import and protein homeostasis (grouped under the pathway of oxidative phosphorylation biogenesis and regulation); the proteins that are involved in nucleotide transport and synthesis; and proteins that are involved in the control of membrane composition and dynamics. The pathways in this figure are adapted from refs 47 and 94.

some mtDNA disorders (such as LHON) are homoplasmic, others (such as MELAS) are heteroplasmic. Cells contain a mixture of wild-type and mutant mtDNA molecules in heteroplasmic disorders, and disease expression only occurs when the mutant mtDNA load exceeds a threshold. Stochastic segregation of mtDNA molecules during development can therefore cause variable tissue expression of disease.

Mutations in nuclear genes

In 1995, the first underlying nuclear gene mutation of a mitochondrial disorder was identified when the Munnich and Rötig groups reported mutations in *SDHA* in a patient with complex II deficiency and Leigh syndrome⁴⁶. Sequencing of the human genome, combined with characterization of the mitochondrial proteome⁴⁷, has propelled the discovery of nuclear disease genes and pathways that are required for respiratory chain assembly and function. Based on recent compilations^{47,48} and disease databases (<http://omim.org>), there are now over 110 nuclear genes that are known to be mutated in respiratory chain disease, and this number is rapidly expanding. The genes that have been identified so far can be organized into five broad pathways involved in the expression, assembly and activity of the oxidative phosphorylation system (Fig. 4 and Table 1).

Mutations in the nuclear-encoded tRNA synthetases underscore the phenotypic heterogeneity that can be associated with the same pathway. Although all of these gene products facilitate translation of 13 mtDNA-encoded respiratory chain proteins, their clinical presentations are quite distinct⁴⁸. For example, mutations in *EARS2* present as leukoencephalopathy and high cerebrospinal fluid lactate, *YARS2* as myopathy and sideroblastic anaemia, *HARS2* as ovarian failure, *AARS2* as hypertrophic cardiomyopathy and *SARS2* as pulmonary hypertension and renal failure. Moreover, these phenotypes do not match those arising from mutations in the corresponding tRNAs that are encoded by the mitochondrial genome. Mitochondrial heterogeneity — perhaps at the level of the nuclear-encoded mitochondrial ribosome⁷ — or moonlighting roles of tRNA synthetases⁴⁹ may underlie the tissue-specific pathology.

Although most of the nuclear genes that underlie respiratory chain disease encode mitochondrial proteins, the small subset that do not provides valuable insight into the cross-talk between the organelle and the rest of the cell. For example, as already mentioned, mitochondria are reliant on some cytosolic pathways for proper nucleotide homeostasis, and mutations in the genes that encode ribonucleotide reductase, *RRM2B* (ref. 12), and thymidine phosphorylase, *TYMP* (ref. 50), disrupt mtDNA maintenance. Mutations in *WFS1*, which encodes an endoplasmic-reticulum resident protein, cause Wolfram syndrome that is characterized by deafness, diabetes mellitus, diabetes insipidus and optic atrophy. Although not formally classified as a mitochondrial disorder, the phenotypic overlap and presence of mtDNA deletions in some patients has led to speculation that *WFS1* mediates interactions between the endoplasmic reticulum and mitochondria⁵¹. This hypothesis is supported by the observation in *Saccharomyces cerevisiae* that a component of the endoplasmic-reticulum–mitochondria encounter structure⁵², *MMM1*, co-localizes with mtDNA nucleoids and has a role in mtDNA stability⁵³. We anticipate that research into the genetics of mitochondrial disorders will continue to reveal unexpected connections, either physical or functional, between mitochondria and the rest of the cell.

Environmental modifiers

Environmental factors can influence the course of genetic mitochondrial disorders, and even phenocopy them. For example, tobacco use and heavy consumption of alcohol are risk factors for loss of vision in LHON⁵⁴. Exposure to toxic substances can produce pathology that resembles LHON, as was seen in an epidemic of blindness that occurred in Cuba in the early 1990s. The cause was ultimately

Table 1 | Products of genes that are known to be mutated in respiratory chain disorders grouped by pathway

Oxidative phosphorylation subunits	mtDNA maintenance and expression	Oxidative phosphorylation biogenesis and regulation	Nucleotide transport and synthesis	Membrane dynamics and composition
Nuclear encoded				
Complex I: NDUFA1, NDUFA2, NDUFA9, NDUFA10, NDUFA11, NDUFA12, NDUFB3, NDUFB9, NDUFS1, NDUFS2, NDUFS3, NDUFS4, NDUFS6, NDUFS7, NDUFS8, NDUFV1, NDUFV2 Complex II: SDHA, SDHB, SDHC, SDHD Complex III: UQCRCB, UQCRCQ Complex IV: COX4I2, COX6B1 Complex V: ATP5E	TWINKLE, MTFMT, GFM1, LRPPRC, MPV17, MRPS16, MRPS22, POLG, POLG2, TRMU, TSFM, TUFM, C12orf65, MTPAP, MRPL3, SARS2, YARS2, HARS2, MARS2, AARS2, RARS2, EARS2, DARS2, TACO1, MTO1, RMND1, PNPT1, PUS1	Complex I: NDUFAF1, NDUFAF2, NDUFAF3, NDUFAF4, NDUFAF5, NDUFAF6, ACAD9, FOXRED1, NUBPL Complex II: SDHAF1, SDHAF2 Complex III: BCS1L, HCCS, TTC19 Complex IV: COX10, COX15, ETHE1, FASTKD2, SCO1, SCO2, SURF1, COX14, COA5 Complex V: ATPAF2, TMEM70 Fe-S: ABCB7, FXN, ISCU, NFU1, BOLA3, GLRX5 Other: DNAJC19, GFER, HSPD1, SPG7, TIMM8A, AIFM1, AFG3L2	DGUOK, RRM2B, SLC25A3, ANT1, SUCLA2, SUCLG1, TK2, TYMP	ADCK3, AGK, COQ2, COQ6, COQ9, DRP1, MFN2, OPA1, PDSS1, PDSS2, TAZ, SERAC1
mtDNA encoded				
Complex I: ND1, ND2, ND3, ND4, ND4L, ND5, ND6 Complex III: CYTB Complex IV: COX1, COX2 Complex V: ATP6, ATP8	12S rRNA, tRNA ^{Tyr} , tRNA ^{Trp} , tRNA ^{Val} , tRNA ^{Thr} , tRNA ^{Ser1} , tRNA ^{Ser2} , tRNA ^{Arg} , tRNA ^{Gln} , tRNA ^{Pro} , tRNA ^{Asn} , tRNA ^{Met} , tRNA ^{Leu1} , tRNA ^{Leu2} , tRNA ^{Lys} , tRNA ^{Ala} , tRNA ^{His} , tRNA ^{Gly} , tRNA ^{Phe} , tRNA ^{Glu} , tRNA ^{Asp} , tRNA ^{Cys} , tRNA ^{Ala}			

List of gene products was generated through synthesis of existing compilations of genes known to be mutated in respiratory-chain disease^{47,48}, as well as review of the literature.

found to be widespread folate deficiency combined with methanol toxicity from homemade rum⁵⁵. Formate, a by-product of methanol metabolism, accumulates in the setting of folate deficiency and causes inhibition of complex IV.

Certain medications have long been known to have toxic effects on mitochondrial function. Owing to the bacterial origins of the mitochondrial ribosome, mtDNA translation can be adversely affected by antibiotics such as aminoglycosides, which can cause sensorineural deafness when administered at high doses. Individuals with certain mtDNA mutations in the 12S ribosomal rRNA — estimated to have a population prevalence of 0.19% — are predisposed to deafness from aminoglycosides, and can experience hearing loss even with short exposure to the recommended doses^{56,57}. The viral origins of the mitochondrial DNA polymerase make it susceptible to nucleoside analogue antivirals (including fialuridine). This class of drugs is used commonly for HIV, and can cause side effects such as lactic acidosis through mtDNA depletion⁵⁸.

Microorganisms are an emerging class of environmental modifiers, ranging from gut microbiota to viruses. Gut bacteria are drivers of disease progression in ethylmalonic encephalopathy. The causal gene, *ETHE1*, encodes an enzyme that detoxifies sulphur compounds, which are released by gut microbiota, and its loss results in H₂S accumulation, with subsequent inhibition of complex IV and short-chain acyl-CoA dehydrogenase. Treatment with metronidazole (an antibiotic that reduces gut microbial content) and *N*-acetylcysteine (which promotes glutathione-mediated detoxification of H₂S) results in clinical improvement⁵⁹. Although a role for viral infections in the course of mitochondrial disorders has not been identified, experimental studies indicate that some viruses, such as HIV⁶⁰, can modulate complex I activity.

Unsolved cases of mitochondrial disease

Over the past 25 years, most of the mitochondrial disease genes have been identified in familial forms of disease, in which it is possible to follow the segregation of highly penetrant causal alleles. In our experience at Massachusetts General Hospital, we have found that less than 25% of patients with clinical and biochemical evidence of mitochondrial disease have strong evidence of an affected first- or second-degree relative. Several recent studies have applied exome sequencing to establish molecular diagnoses in singleton cases^{61,62}. However, the success rate is lower than anticipated. Our experience has shown that in biochemically proven, severe cases in infants exome sequencing should achieve a diagnosis in about half of cases⁶²;

however, the success rate is projected to be much lower in milder cases of disease, or those with adult onset.

How can we explain these unsolved cases? Most exome studies of singleton cases have been powered to identify recessive mutations in mitochondrial proteins. It is possible that these unsolved cases are a result of dominant-acting, subtle recessive or regulatory mutations with incomplete penetrance, all of which are difficult to identify over the background of polymorphisms. Alternatively, these cases could be due to mutations in regions of DNA that are not targeted for sequencing. Some cases could also have purely environmental causes, as with the Cuban blindness epidemic. A tantalizing possibility is that a subset of these unsolved cases is due to complex genetic inheritance, as a result of interactions between gene variants that each have weak or synergistic effects, known as synergistic heterozygosity⁶³. Targeted exome sequencing studies have reported that healthy controls will typically carry a burden of about 15–20 heterozygous, loss-of-function protein alleles within their mitochondrial proteomes⁶². This high burden of deleterious alleles is probably tolerated because of the robustness of mitochondrial networks. However, it is possible that mutations that affect multiple genes, operating in the same or parallel pathways, may conspire to yield pathology. Notably, there is suggestive evidence that supports synergistic interactions between mtDNA and nuclear DNA variants⁶⁴. Defining the genetic architecture of the large number of unsolved sporadic cases of mitochondrial disease represents the next major challenge of mitochondrial genetics.

Mitochondrial ripples and responses

Despite remarkable progress in defining the genes and environmental triggers that underlie mitochondrial disease, their pathogenesis remains almost a complete enigma. Many medical textbooks offer the oversimplified explanation that disease manifests in tissues with the highest ATP demand, or because of oxidative damage. Although these factors probably contribute to disease, the picture is much more complicated. Cellular models of disease indicate that there is a remarkable capacity for preservation of ATP production through enhanced glycolysis⁶⁵, and, in animal models of respiratory chain dysfunction, pathology can develop without a major increase in oxidative damage^{66,67}. Furthermore, the fact that inhibition of respiration can be tolerated, and even beneficial in certain settings (Box 1), indicates that the consequences of respiratory chain lesions are not uniformly bad, and suggests the involvement of nonlinear modes of pathogenesis and threshold effects.

BOX 1

Can inhibition of the respiratory chain be beneficial?

Although respiratory chain inhibition is often viewed as undesirable, several observations have suggested that the consequences can be beneficial in certain situations. Metformin, the most commonly prescribed oral medication for type 2 diabetes mellitus, inhibits complex I. This bioenergetic effect is thought to contribute to metformin's inhibition of hepatic gluconeogenesis⁹⁵. Lactic acidosis is a known — although infrequent — side effect of metformin treatment, highlighting the fine balance between therapeutic and toxic effects of respiratory chain inhibition. Pretreatment with small-molecule inhibitors of the respiratory chain can protect organs, such as the brain and

heart, against ischaemia–reperfusion injury⁹⁶. Finally, data from studies in both the roundworm *Caenorhabditis elegans*⁹⁷ and the fruitfly *Drosophila melanogaster*⁹⁸ indicate that RNA-interference-mediated knockdown of several components of the respiratory chain extends lifespan. The general theme that emerges from all of these observations is that the cellular and organismal consequences of modest respiratory chain inhibition can be used for therapeutic purposes — a concept that some have termed mitochondrial hormesis. We suspect that therapeutic effects occur when the balance of responses to a respiratory chain lesion weighs in favour of homeostasis as opposed to pathogenesis.

To systematically understand pathogenesis, it is convenient to consider the proximal consequences of respiratory chain lesions — or mitochondrial ripples — and the secondary cellular responses they evoke. Model-organism studies provide several examples of such ripple–response cascades. The ‘retrograde’ response in *S. cerevisiae* is a defined transcriptional response that allows the survival of that organism in the setting of an impaired respiratory chain, mainly by compensating for incomplete TCA-cycle function to maintain production of metabolites such as glutamate⁶⁸. The precise ripples that trigger the yeast retrograde response remain unclear, although loss of the PMF is thought to have a key role⁶⁹. In the fruitfly *Drosophila melanogaster*, respiratory chain impairment can trigger blockade of the G1–S transition of the cell cycle through two possible ripple–response pathways: mutation of a complex-I subunit leads to ROS-mediated activation of JNK signalling⁷⁰, whereas diminished ATP production as a consequence of a complex-IV mutation causes AMPK activation⁷¹. Studies in the roundworm *Caenorhabditis elegans* have defined a mitochondrial unfolded protein response (UPR^{mt}) that is activated by disturbed protein homeostasis resulting from insults to the respiratory chain⁷². The UPR^{mt} comprises increased expression of mitochondrial chaperones, and, notably, mutations in this pathway can give rise to several neurodegenerative disorders⁷³. The lessons that have emerged from model-organism studies are that a broad range of ripple–response pairs can result from respiratory chain lesions, and that these pairs confer remarkable tolerance to such insults. However, some of these may be adaptive on short timescales and pathology-inducing over longer timescales.

Metabolomics studies have attempted to systemically catalogue biochemical ripples that emanate from respiratory chain inhibition. One study characterized the effect of small-molecule inhibitors of the respiratory chain on metabolite flux in cultured cells⁷⁴ and correlated these changes to patient plasma measurements. Specific blockade of complex III resulted in decreased production of uridine, which was consistent with the long-standing observation that respiratory chain-deficient cells are uridine auxotrophs⁷⁵. Lactate secretion highlighted that the induction of the lactate dehydrogenase reaction is a means to support glycolytic ATP production and consume NADH to maintain redox cofactor balance. Another study reported that reverse flux through the TCA cycle, specifically reductive carboxylation of α -ketoglutarate to isocitrate, is a response to respiratory chain inhibition⁷⁶.

Ideally, investigation of ripple–response cascades should explain the end pathology. One promising cascade involves calcium. Mitochondria have a major role in shaping the cytosolic calcium concentrations through uptake with a uniporter^{77–79}, which is dependent on an intact PMF. Several studies have converged on the mechanism that an increase in cytosolic calcium, secondary to loss of

mitochondrial PMF, is a key signalling intermediate in response to respiratory chain lesions through activation of calcium-sensitive signalling factors (such as calcineurin and calcium/calmodulin-dependent protein kinase IV (CaMKIV)^{80,81}). We hypothesize that activation of calcium-dependent signalling may in fact help to explain several pathological hallmarks (Fig. 1). For example, activation of CaMKIV can induce mitochondrial biogenesis⁸², potentially contributing to the finding of ‘ragged red fibres’ that are often seen on skeletal-muscle biopsy. Activation of calcineurin has also been found to induce hypertrophic cardiomyopathy⁸³. Altered calcium dynamics in gastrointestinal interstitial cells of Cajal compromise their pacemaking activity⁸⁴, and this may contribute to intestinal pseudo-obstruction (Fig. 1b). However, not all respiratory chain mutations disrupt calcium homeostasis⁸⁵. Moreover, mitochondrial calcium buffering can vary across tissues⁸⁶, and this may shape the pattern of tissue expression.

Ripple–response cascades can flow outside of cells, leading to non-cell-autonomous effects. Lactic acidosis is the best known example, and can impair the function of multiple organs by reducing serum pH. Fibroblast growth factor 21, a hormone that mediates aspects of the starvation response, is released from the skeletal muscle of patients with mitochondrial myopathy⁸⁷. This has led to the intriguing hypothesis that it may drive systemic metabolic pathology. Recent evidence has pointed towards a crucial role for mitochondria in the regulation of innate immunity⁸⁸; however, its role in mitochondrial disorders is largely unexplored.

Future prospects and challenges

Understanding the pathogenesis of mitochondrial disorders is an exciting new frontier, with many opportunities. First, as a group, these disorders affect at least 1 in 5,000 individuals, and there are no proven therapies⁸⁹. There is, therefore, a special opportunity to develop therapeutics for patients who otherwise have distressingly few options. Second, these disorders are a continuous source of insight into basic cell biology, and much of what we know about respiratory chain assembly and compartmentalization of metabolism has come directly from studying them. Finally, there is great interest in the role of mitochondria in many common human diseases, which is fuelled by the observation that many common, degenerative disorders are associated with a quantitative decline in mitochondrial activity. Distinguishing cause from correlation, however, is challenging and mitochondrial disorders may be valuable tools for clarifying the organelle's role in common disease. Capitalizing on these opportunities will require that we fully decipher the pathogenesis of these orphan diseases.

We anticipate that two complementary lines of investigation, cell- and patient-based, will be required to meet this challenge. First, it will be essential to characterize the network-level properties

of mitochondria in cell-based studies. As a result of classic biochemical studies and (more recent) proteomic studies of isolated mitochondria, we have a reasonable knowledge of the organelle's protein inventory and reaction repertoire, as well as a theoretical framework for understanding regulatory control. Moving forward, it will be crucial to experimentally map interactions among all of its components, both at a physical and genetic level, across different cell types. Initial progress has already been made in this area, with the recent production of a genetic interaction map, focusing on *S. cerevisiae* mitochondria⁹⁰. The organelle's rich evolutionary history and transcriptional regulation will facilitate computational identification of modules of functionally interacting genes⁹¹. These network-level maps will reveal the homeostatic mechanisms that buffer against environmental or genetic insults. Second, in parallel, it will be crucial to catalogue genotypes and phenotypes from individual patients throughout the world. Next-generation sequencing is already facilitating sequencing of patient genomes, while Internet- and wireless-enabled technologies will yield in-depth phenotypes with unprecedented temporal resolution. Because the individual mitochondrial disorders are so rare and diverse, we anticipate that an open-source, collaborative model — in which patients and their doctors are partners for biomedical research — will be required so that data can be aggregated and shared for discovery. Metabolic profiles, which are obtained from perturbed cells grown in culture as well as from human plasma⁷⁴, may represent the key ingredient for connecting cell-level and patient-level data into predictive models of pathogenesis.

Identifying a link between genotype and phenotype is a challenge for all diseases. But features unique to mitochondria — their relatively well-studied biochemistry, a near-complete protein parts list, an ability to study mitochondria in isolation or *in situ*, the large number of monogenic disorders, and an active and engaged patient community — provide key advantages that promise to place this organelle at the forefront of the burgeoning field of medical systems biology.

Studies of orphan mitochondrial disorders will be pivotal in solving one of the most important problems in fundamental metabolism: the logic of compartmentalization. Why have certain pathways persisted within mitochondria, whereas others have been duplicated or fully relocated to the cytosol? Mitochondria have retained a bacterial type-II fatty-acid synthesis pathway, as well as duplicate versions of tetrahydrofolate-dependent one-carbon metabolism and the gluconeogenic enzyme phosphoenolpyruvate carboxykinase. What advantage is conferred by the gain or loss of such pathways? We anticipate that studies of the rare mitochondrial disorders will shed light on the *in vivo* relevance of these pathways, and the circumstances in which they are operative. New ways of measuring compartment-specific metabolism, both in cultured cells and *in vivo*, will be required to drive this field forward.

We can be optimistic that understanding mitochondrial pathogenesis will enable the development of new therapeutics for these devastating disorders, some of which may also be useful for the treatment of common diseases. An example of such repurposing can be seen in the history of dichloroacetate, a small molecule that reduces lactic acid production by stimulating pyruvate dehydrogenase. Dichloroacetate was tested for the treatment of lactic acidosis in mitochondrial disorders but lacked clinical efficacy and resulted in toxic side effects⁹². Although it was largely unsuccessful for use in the treatment of mitochondrial disorders, it has since been repurposed and is now being tested in clinical trials for some cancers. Many cancers have an increased reliance on aerobic glycolysis with concomitant lactate production — the so-called Warburg effect. Dichloroacetate is being used to reverse this metabolic hallmark, with promising early results in human trials⁹³. We anticipate that this pattern will continue, and that drug development for mitochondrial disorders will ultimately prove beneficial for patients with these rare disorders, as well as those with more common disease. ■

- Ernster, L., Ikkos, D. & Luft, R. Enzymic activities of human skeletal muscle mitochondria: a tool in clinical metabolic research. *Nature* **184**, 1851–1854 (1959).
This paper reports a fascinating case of euthyroid hypermetabolism, which is now regarded as the first case of a biochemically proven mitochondrial disease.
- DiMauro, S. *et al.* Luft's disease. Further biochemical and ultrastructural studies of skeletal muscle in the second case. *J. Neurol. Sci.* **27**, 217–232 (1976).
- Skladal, D., Halliday, J. & Thorburn, D. R. Minimum birth prevalence of mitochondrial respiratory chain disorders in children. *Brain* **126**, 1905–1912 (2003).
- Munnich, A., Rotig, A., Cormier-Daire, V. & Rustin, P. in *Scriver's Online Metabolic and Molecular Basis of Inherited Disease* (eds Valle, D. *et al.*) Ch. 99 <http://dx.doi.org/10.1036/ommbid.127> (McGraw-Hill, 2006)
- Shoffner, J. in *Scriver's Online Metabolic and Molecular Basis of Inherited Disease* (eds Valle, D. *et al.*) Ch. 104 <http://dx.doi.org/10.1036/ommbid.127> (McGraw-Hill, 2006).
- McKenzie, R. *et al.* Hepatic failure and lactic acidosis due to fialuridine (FIAU), an investigational nucleoside analogue for chronic hepatitis B. *N. Engl. J. Med.* **333**, 1099–1105 (1995).
- Pagliarini, D. J. *et al.* A mitochondrial protein compendium elucidates complex I disease biology. *Cell* **134**, 112–123 (2008).
This paper reports an accurate inventory of the mammalian mitochondrial proteome, called MitoCarta, consisting of about 1,100 proteins.
- Andersson, S. G. *et al.* The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* **396**, 133–140 (1998).
This paper reports the complete genome sequence of *Rickettsia prowazekii*, providing support for an α -proteobacterial ancestor of human mitochondria.
- Szklarczyk, R. & Huynen, M. A. Mosaic origin of the mitochondrial proteome. *Proteomics* **10**, 4012–4024 (2010).
- Sharma, M. R. *et al.* Structure of the mammalian mitochondrial ribosome reveals an expanded functional role for its component proteins. *Cell* **115**, 97–108 (2003).
- Shutt, T. E. & Gray, M. W. Bacteriophage origins of mitochondrial replication and transcription proteins. *Trends Genet.* **22**, 90–95 (2006).
- Bourdon, A. *et al.* Mutation of *RRM2B*, encoding p53-controlled ribonucleotide reductase (p53R2), causes severe mitochondrial DNA depletion. *Nature Genet.* **39**, 776–780 (2007).
- Tibbetts, A. S. & Appling, D. R. Compartmentalization of mammalian folate-mediated one-carbon metabolism. *Annu. Rev. Nutr.* **30**, 57–81 (2010).
- Jain, M. *et al.* Metabolite profiling identifies a key role for glycine in rapid cancer cell proliferation. *Science* **336**, 1040–1044 (2012).
- Hackenbrock, C. R., Chazotte, B. & Gupte, S. S. The random collision model and a critical assessment of diffusion and collision in mitochondrial electron transport. *J. Bioenerg. Biomembr.* **18**, 331–368 (1986).
- Chance, B. & Williams, G. R. A method for the localization of sites for oxidative phosphorylation. *Nature* **176**, 250–254 (1955).
- Schagger, H. & Pfeiffer, K. The ratio of oxidative phosphorylation complexes I–V in bovine heart mitochondria and the composition of respiratory chain supercomplexes. *J. Biol. Chem.* **276**, 37861–37867 (2001).
- Schagger, H. & von Jagow, G. Blue native electrophoresis for isolation of membrane protein complexes in enzymatically active form. *Anal. Biochem.* **199**, 223–231 (1991).
- Cruciat, C. M., Brunner, S., Baumann, F., Neupert, W. & Stuart, R. A. The cytochrome *bc₁* and cytochrome *c* oxidase complexes associate to form a single supracomplex in yeast mitochondria. *J. Biol. Chem.* **275**, 18093–18098 (2000).
- Shoubridge, E. A. Supersizing the mitochondrial respiratory chain. *Cell Metab.* **15**, 271–272 (2012).
- Efremov, R. G., Baradaran, R. & Sazanov, L. A. The architecture of respiratory complex I. *Nature* **465**, 441–445 (2010).
- Hunte, C., Zickermann, V. & Brandt, U. Functional modules and structural basis of conformational coupling in mitochondrial complex I. *Science* **329**, 448–451 (2010).
- Runswick, M. J., Fearnley, I. M., Skehel, J. M. & Walker, J. E. Presence of an acyl carrier protein in NADH:ubiquinone oxidoreductase from bovine heart mitochondria. *FEBS Lett.* **286**, 121–124 (1991).
- Nouws, J. *et al.* Acyl-CoA dehydrogenase 9 is required for the biogenesis of oxidative phosphorylation complex I. *Cell Metab.* **12**, 283–294 (2010).
- Cogswell, A. M., Stevens, R. J. & Hood, D. A. Properties of skeletal muscle mitochondria isolated from subsarcolemmal and intermyofibrillar regions. *Am. J. Physiol.* **264**, C383–C389 (1993).
- Fannin, S. W., Lesnfsky, E. J., Slabe, T. J., Hassan, M. O. & Hoppel, C. L. Aging selectively decreases oxidative capacity in rat heart intermyofibrillar mitochondria. *Arch. Biochem. Biophys.* **372**, 399–407 (1999).
- Allen, J. F. The function of genomes in bioenergetic organelles. *Phil. Trans. R. Soc. Lond. B* **358**, 19–37 (2003).
- Johnson, D. T., Harris, R. A., Blair, P. V. & Balaban, R. S. Functional consequences of mitochondrial proteome heterogeneity. *Am. J. Physiol. Cell Physiol.* **292**, C698–C707 (2007).
- Pierron, D. *et al.* Cytochrome *c* oxidase: evolution of control via nuclear subunit addition. *Biochim. Biophys. Acta* **1817**, 590–597 (2012).
- Scarpulla, R. C. Metabolic control of mitochondrial biogenesis through the PGC-1 family regulatory network. *Biochim. Biophys. Acta* **1813**, 1269–1278 (2011).

31. Spiegelman, B. M. Transcriptional control of mitochondrial energy metabolism through the PGC1 coactivators. *Novartis Found. Symp.* **287**, 60–63 (2007).
32. MacAskill, A. F. & Kittler, J. T. Control of mitochondrial transport and localization in neurons. *Trends Cell Biol.* **20**, 102–112 (2010).
33. Chan, D. C. Fusion and fission: interlinked processes critical for mitochondrial health. *Annu. Rev. Genet.* <http://dx.doi.org/10.1146/annurev-genet-110410-132529> (29 August, 2012).
34. Youle, R. J. & Narendra, D. P. Mechanisms of mitophagy. *Nature Rev. Mol. Cell Biol.* **12**, 9–14 (2011).
35. Balaban, R. S., Kantor, H. L., Katz, L. A. & Briggs, R. W. Relation between work and phosphate metabolite in the *in vivo* paced mammalian heart. *Science* **232**, 1121–1123 (1986).
36. Chance, B. & Williams, G. R. Respiratory enzymes in oxidative phosphorylation. III. The steady state. *J. Biol. Chem.* **217**, 409–427 (1955).
This classic paper introduced the ‘respiratory states’ of isolated mitochondria, providing a conceptual framework for studying mitochondrial energetics.
37. Glancy, B. & Balaban, R. S. Role of mitochondrial Ca^{2+} in the regulation of cellular energetics. *Biochemistry* **51**, 2959–2973 (2012).
38. Kacser, H. & Burns, J. A. The control of flux. *Symp. Soc. Exp. Biol.* **27**, 65–104 (1973).
39. Heinrich, R. & Rapoport, T. A. A linear steady-state treatment of enzymatic chains. General properties, control and effector strength. *Eur. J. Biochem.* **42**, 89–95 (1974).
40. Groen, A. K., Wanders, R. J., Westerhoff, H. V., van der Meer, R. & Tager, J. M. Quantification of the contribution of various steps to the control of mitochondrial respiration. *J. Biol. Chem.* **257**, 2754–2757 (1982).
41. Hartwell, L. Genetics. Robust interactions. *Science* **303**, 774–775 (2004).
42. Rossignol, R., Malgat, M., Mazat, J. P. & Letellier, T. Threshold effect and tissue specificity. Implication for mitochondrial cytopathies. *J. Biol. Chem.* **274**, 33426–33432 (1999).
43. Anderson, S. *et al.* Sequence and organization of the human mitochondrial genome. *Nature* **290**, 457–465 (1981).
This landmark publication reports the sequence and annotation of the human mitochondrial genome.
44. Wallace, D. C. *et al.* Mitochondrial DNA mutation associated with Leber’s hereditary optic neuropathy. *Science* **242**, 1427–1430 (1988).
45. Holt, I. J., Harding, A. E. & Morgan-Hughes, J. A. Deletions of muscle mitochondrial DNA in patients with mitochondrial myopathies. *Nature* **331**, 717–719 (1988).
References 44 and 45 report the first disease-causing mutations in the mitochondrial genome (mtDNA).
46. Bourgeron, T. *et al.* Mutation of a nuclear succinate dehydrogenase gene results in mitochondrial respiratory chain deficiency. *Nature Genet.* **11**, 144–149 (1995).
This paper reports the first nuclear gene mutation that gives rise to a mitochondrial respiratory chain disorder.
47. Calvo, S. E. & Mootha, V. K. The mitochondrial proteome and human disease. *Annu. Rev. Genomics Hum. Genet.* **11**, 25–44 (2010).
48. Koopman, W. J., Willems, P. H. & Smeitink, J. A. Monogenic mitochondrial disorders. *N. Engl. J. Med.* **366**, 1132–1141 (2012).
49. Park, S. G., Schimmel, P. & Kim, S. Aminoacyl tRNA synthetases and their connections to disease. *Proc. Natl Acad. Sci. USA* **105**, 11043–11049 (2008).
50. Nishino, I., Spinazzola, A. & Hirano, M. Thymidine phosphorylase gene mutations in MNGIE, a human mitochondrial disorder. *Science* **283**, 689–692 (1999).
51. Lieber, D. S. *et al.* Atypical case of Wolfram syndrome revealed through targeted exome sequencing in a patient with suspected mitochondrial disease. *BMC Med. Genet.* **13**, 3 (2012).
52. Kornmann, B. *et al.* An ER-mitochondria tethering complex revealed by a synthetic biology screen. *Science* **325**, 477–481 (2009).
53. Hobbs, A. E., Srinivasan, M., McCaffery, J. M. & Jensen, R. E. Mmm1p, a mitochondrial outer membrane protein, is connected to mitochondrial DNA (mtDNA) nucleoids and required for mtDNA stability. *J. Cell Biol.* **152**, 401–410 (2001).
54. Kirkman, M. A. *et al.* Gene–environment interactions in Leber hereditary optic neuropathy. *Brain* **132**, 2317–2326 (2009).
55. Sadun, A. Acquired mitochondrial impairment as a cause of optic nerve disease. *Trans. Am. Ophthalmol. Soc.* **96**, 881–923 (1998).
56. Bitner-Glindzic, M. *et al.* Prevalence of mitochondrial 1555A>G mutation in European children. *N. Engl. J. Med.* **360**, 640–642 (2009).
57. Prezant, T. R. *et al.* Mitochondrial ribosomal RNA mutation associated with both antibiotic-induced and non-syndromic deafness. *Nature Genet.* **4**, 289–294 (1993).
58. Cote, H. C. *et al.* Changes in mitochondrial DNA as a marker of nucleoside toxicity in HIV-infected patients. *N. Engl. J. Med.* **346**, 811–820 (2002).
59. Viscomi, C. *et al.* Combined treatment with oral metronidazole and N-acetylcysteine is effective in ethylmalonic encephalopathy. *Nature Med.* **16**, 869–871 (2010).
60. Reeves, M. B., Davies, A. A., McSharry, B. P., Wilkinson, G. W. & Sinclair, J. H. Complex I binding by a virally encoded RNA regulates mitochondria-induced cell death. *Science* **316**, 1345–1348 (2007).
61. Vasta, V., Merritt, J. L. 2nd, Saneto, R. P. & Hahn, S. H. Next-generation sequencing for mitochondrial diseases reveals wide diagnostic spectrum. *Pediatr. Int.* **54**, 585–601 (2012).
62. Calvo, S. E. *et al.* Molecular diagnosis of infantile mitochondrial disease with targeted next-generation sequencing. *Sci. Transl. Med.* **4**, 118ra110 (2012).
63. Vockley, J., Rinaldo, P., Bennett, M. J., Matern, D. & Vladutiu, G. D. Synergistic heterozygosity: disease resulting from multiple partial defects in one or more metabolic pathways. *Mol. Genet. Metab.* **71**, 10–18 (2000).
64. Guan, M. X. *et al.* Mutation in *TRMU* related to transfer RNA modification modulates the phenotypic expression of the deafness-associated mitochondrial 12S ribosomal RNA mutations. *Am. J. Hum. Genet.* **79**, 291–302 (2006).
65. von Kleist-Retzow, J. C. *et al.* Impaired mitochondrial Ca^{2+} homeostasis in respiratory chain-deficient cells but efficient compensation of energetic disadvantage by enhanced anaerobic glycolysis due to low ATP steady state levels. *Exp. Cell Res.* **313**, 3076–3089 (2007).
66. Trifunovic, A. *et al.* Somatic mtDNA mutations cause aging phenotypes without affecting reactive oxygen species production. *Proc. Natl Acad. Sci. USA* **102**, 17993–17998 (2005).
67. Kujoth, G. C. *et al.* Mitochondrial DNA mutations, oxidative stress, and apoptosis in mammalian aging. *Science* **309**, 481–484 (2005).
68. Butow, R. A. & Avadhani, N. G. Mitochondrial signaling: the retrograde response. *Mol. Cell* **14**, 1–15 (2004).
69. Miceli, M. V., Jiang, J. C., Tiwari, A., Rodriguez-Quinones, J. F. & Jazwinski, S. M. Loss of mitochondrial membrane potential triggers the retrograde response extending yeast replicative lifespan. *Front. Genet.* **2**, 102 (2011).
70. Owusu-Ansah, E., Yavari, A., Mandal, S. & Banerjee, U. Distinct mitochondrial retrograde signals control the G1-S cell cycle checkpoint. *Nature Genet.* **40**, 356–361 (2008).
71. Mandal, S., Gupta, P., Owusu-Ansah, E. & Banerjee, U. Mitochondrial regulation of cell cycle progression during development as revealed by the *tenured* mutation in *Drosophila*. *Dev. Cell* **9**, 843–854 (2005).
72. Haynes, C. M. & Ron, D. The mitochondrial UPR — protecting organelle protein homeostasis. *J. Cell Sci.* **123**, 3849–3855 (2010).
73. Rugar, E. I. & Langer, T. Mitochondrial quality control: a matter of life and death for neurons. *EMBO J.* **31**, 1336–1349 (2012).
74. Shaham, O. *et al.* A plasma signature of human mitochondrial disease revealed through metabolic profiling of spent media from cultured muscle cells. *Proc. Natl Acad. Sci. USA* **107**, 1571–1575 (2010).
This paper reports the application of metabolite profiling to systematically characterize the biochemical ripples that ensue from defined lesions to the mitochondrial respiratory chain, some of which are also mirrored in the plasma of patients with mitochondrial disease.
75. Morais, R., Guertin, D. & Kornblatt, J. A. On the contribution of the mitochondrial genome to the growth of Chinese hamster embryo cells in culture. *Can. J. Biochem.* **60**, 290–294 (1982).
76. Mullen, A. R. *et al.* Reductive carboxylation supports growth in tumour cells with defective mitochondria. *Nature* **481**, 385–388 (2012).
77. Perocchi, F. *et al.* *MICU1* encodes a mitochondrial EF hand protein required for Ca^{2+} uptake. *Nature* **467**, 291–296 (2010).
78. Baughman, J. M. *et al.* Integrative genomics identifies MCU as an essential component of the mitochondrial calcium uniporter. *Nature* **476**, 341–345 (2011).
79. De Stefani, D., Raffaello, A., Teardo, E., Szabo, I. & Rizzuto, R. A forty-kilodalton protein of the inner membrane is the mitochondrial calcium uniporter. *Nature* **476**, 336–340 (2011).
80. Biswas, G. *et al.* Retrograde Ca^{2+} signaling in C2C12 skeletal myocytes in response to mitochondrial genetic and metabolic stress: a novel mode of inter-organelle crosstalk. *EMBO J.* **18**, 522–533 (1999).
81. Arnould, T. *et al.* CREB activation induced by mitochondrial dysfunction is a new signaling pathway that impairs cell proliferation. *EMBO J.* **21**, 53–63 (2002).
82. Wu, H. *et al.* Regulation of mitochondrial biogenesis in skeletal muscle by CaMK. *Science* **296**, 349–352 (2002).
83. Molken, J. D. *et al.* A calcineurin-dependent transcriptional pathway for cardiac hypertrophy. *Cell* **93**, 215–228 (1998).
84. Ward, S. M. *et al.* Pacemaking in interstitial cells of Cajal depends upon calcium handling by endoplasmic reticulum and mitochondria. *J. Physiol.* **525**, 355–361 (2000).
85. Brini, M. *et al.* A calcium signalling defect in the pathogenesis of a mitochondrial DNA inherited oxidative phosphorylation deficiency. *Nature Med.* **5**, 951–954 (1999).
86. Kaftan, E. J., Xu, T., Abercrombie, R. F. & Hille, B. Mitochondria shape hormonally induced cytoplasmic calcium oscillations and modulate exocytosis. *J. Biol. Chem.* **275**, 25465–25470 (2000).
87. Suomalainen, A. *et al.* FGF-21 as a biomarker for muscle-manifesting mitochondrial respiratory chain deficiencies: a diagnostic study. *Lancet Neurol.* **10**, 806–818 (2011).
88. West, A. P., Shadel, G. S. & Ghosh, S. Mitochondria in innate immune responses. *Nature Rev. Immunol.* **11**, 389–402 (2011).
89. Pfeffer, G., Majumaa, K., Turnbull, D. M., Thorburn, D. & Chinnery, P. F. Treatment for mitochondrial disorders. *Cochrane Database Syst. Rev.* **4**, CD004426 (2012).
90. Hoppins, S. *et al.* A mitochondrial-focused genetic interaction map reveals a scaffold-like complex required for inner membrane organization in mitochondria. *J. Cell Biol.* **195**, 323–340 (2011).

91. Nilsson, R. *et al.* Discovery of genes essential for heme biosynthesis through large-scale gene expression analysis. *Cell Metab.* **10**, 119–130 (2009).
92. Stacpoole, P. W. Why are there no proven therapies for genetic mitochondrial diseases? *Mitochondrion* **11**, 679–685 (2011).
93. Michelakis, E. D. *et al.* Metabolic modulation of glioblastoma with dichloroacetate. *Sci. Transl. Med.* **2**, 31ra34 (2010).
94. Kirby, D. M. & Thorburn, D. R. Approaches to finding the molecular basis of mitochondrial oxidative phosphorylation disorders. *Twin Res. Hum. Genet.* **11**, 395–411 (2008).
95. Owen, M. R., Doran, E. & Halestrap, A. P. Evidence that metformin exerts its anti-diabetic effects through inhibition of complex 1 of the mitochondrial respiratory chain. *Biochem. J.* **348**, 607–614 (2000).
96. Gohil, V. M. *et al.* Nutrient-sensitized screening for drugs that shift energy metabolism from mitochondrial respiration to glycolysis. *Nature Biotechnol.* **28**, 249–255 (2010).
97. Copeland, J. M. *et al.* Extension of *Drosophila* life span by RNAi of the mitochondrial respiratory chain. *Curr. Biol.* **19**, 1591–1598 (2009).
98. Lee, S. S. *et al.* A systematic RNAi screen identifies a critical role for mitochondria in *C. elegans* longevity. *Nature Genet.* **33**, 40–48 (2003).

Acknowledgements We apologize to the many authors whose work we were unable to cite because of space limitations. We offer special thanks to D. Thorburn for careful review of the manuscript and his help with compiling an updated list of disease genes. We are grateful to S. Calvo, M. Jain, E. Rosen, V. Siegel and M. Gray for thoughtful feedback on the manuscript; J.-P. Mazat for providing a figure; M. Fleming, A. Sadun, M. Seidman, R. Mitchell, R. Saneto, D. McGuone and L. Rodriguez for providing clinical images; and G. Perkins and M. Ellisman for providing electron micrographs. We thank the National Institutes of Health for ongoing grant support.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at go.nature.com/8qufjj. Correspondence should be addressed to V.M. (vamsi@hms.harvard.edu).

Metabolic phenotyping in clinical and surgical environments

Jeremy K. Nicholson¹, Elaine Holmes¹, James M. Kinross¹, Ara W. Darzi¹, Zoltan Takats¹ & John C. Lindon¹

Metabolic phenotyping involves the comprehensive analysis of biological fluids or tissue samples. This analysis allows biochemical classification of a person's physiological or pathological states that relate to disease diagnosis or prognosis at the individual level and to disease risk factors at the population level. These approaches are currently being implemented in hospital environments and in regional phenotyping centres worldwide. The ultimate aim of such work is to generate information on patient biology using techniques such as patient stratification to better inform clinicians on factors that will enhance diagnosis or the choice of therapy. There have been many reports of direct applications of metabolic phenotyping in a clinical setting.

Clinical diagnosis, prognosis and treatment selection are increasingly dependent on the use of molecular tools that help to classify diseases and their subtypes, and to define underlying individual variations in patient biology. The application of stratified and new therapeutic approaches that have been optimized through predictive modelling of deep biological information (for example, genetic, metabolic or physiological) on individual patient variation will not only have major health-care benefits, but also inevitably lead to socioeconomic, health-care deployment, regulatory and research changes in the clinic¹. One of the most widely applicable areas for the development of precision medicine relates to the diverse applications of metabolic phenotyping — or metabotyping² — to clinical diagnostics, prognostics and molecular epidemiology. The metabotypes of individuals can be measured from the composition of accessible biofluids or tissues that are sampled in the clinic. Metabotypes vary extensively between individuals and populations, and result from the complex interplay between host genes, lifestyle, diet and gut microbes^{3,4}. Thus, metabotyping has applications in both population-based disease-risk investigation studies and in solving problems related to personalized health care and patient stratification⁴. Hence, the ability to generate metabolic phenotypes from large sample cohorts that have been collected as part of epidemiological studies means that the ensuing enormous statistical power allows the identification of good candidates for metabolic biomarkers of disease risk in different populations (such as predictors of elevated blood pressure in so-called metabolome-wide association studies³). The gene–environment interactions that determine metabotypes are identical to those that determine disease risk in the general population, as well as individual susceptibility to disease and response to treatment. Thus, metabotypes are both statistically and biologically connected to disease risk factors and treatment outcomes, and thereby underpin the value of metabolic analysis in a diverse range of medical scenarios⁴. Metabolic phenotypes have been measured and mapped indirectly for many centuries — mainly unknowingly. For example, the urine wheel was used by physicians to relate the colours, smells and tastes of urine samples to likely diagnoses and treatments⁵. More recently, spectroscopic methods have been applied to generate multivariate profiles of metabolites, mainly using nuclear magnetic resonance (NMR) or mass-spectrometric methods that can measure a wide range of metabolites simultaneously. The data are then analysed using multivariate statistics (Fig. 1 and Box 1).

A number of terms are used to describe the various metabolic-analysis procedures. Metabolomics⁶, for example, essentially describes the metabolic composition of a given sample in terms of metabolite presence and concentration, the metabolome being the multivariate sum of these components. There are about 500 histologically distinct cell types in the human body. Each one of these cell types has specific functions and, consequently, a different gene expression pattern, proteome and metabolome. Cellular metabotypes may overlap within histological specimens, but they interact in space and time through the connecting vascular and lymphatic systems. Humans, therefore, contain more than 500 dynamic cellular metabolomes, as well as those of the individual tissue-specific extracellular fluid compartments (which are compositionally different from their surrounding cells) and the various secretory and excretory biological fluids (Fig. 2). Disease processes and medical treatments also occur over variable time frames, and metabotypes change dynamically with disease and treatment. Thus, the term metabonomics has been used since the late 1990s (ref. 7) to describe the metabolic responses of complex systems to perturbations through time, and how these responses can be mapped using appropriate analytical and statistical techniques. This stimulus could be disease, nutritional changes, drug therapy, genetic modulation or a myriad of other inputs. Specifically, metabonomics addresses such phenotypic changes at the level of small-molecule metabolites, and usually in the context of analysis of body fluids such as urine or blood plasma. The terms metabolomics and metabonomics are widely — and often interchangeably — used, having received about 26,000 and about 10,400 Google Scholar hits, respectively, at the time of writing. Although metabolic profiling has been applied in a wide variety of research fields over the past 30 years — ranging from microbiology to plant and food science, through to animal toxicology and mechanisms of disease — it is the clinical areas that are currently receiving the most attention. Applying high-throughput metabolic technologies to provide new diagnostic biomarkers and to uncover disease mechanisms is an attractive proposition. In this Review, we discuss some of the key and developing areas in clinical metabotyping, and its range of applications in progressing our understanding of human disease processes.

Metabolic analysis of biofluids, cells and tissue

A series of interacting metabolic networks that operate in multiple body compartments gives rise to a continuum of metabolic processes that contribute to the overall metabotype, and includes contributions from diet,

¹Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, South Kensington, London SW7 2AZ, UK.

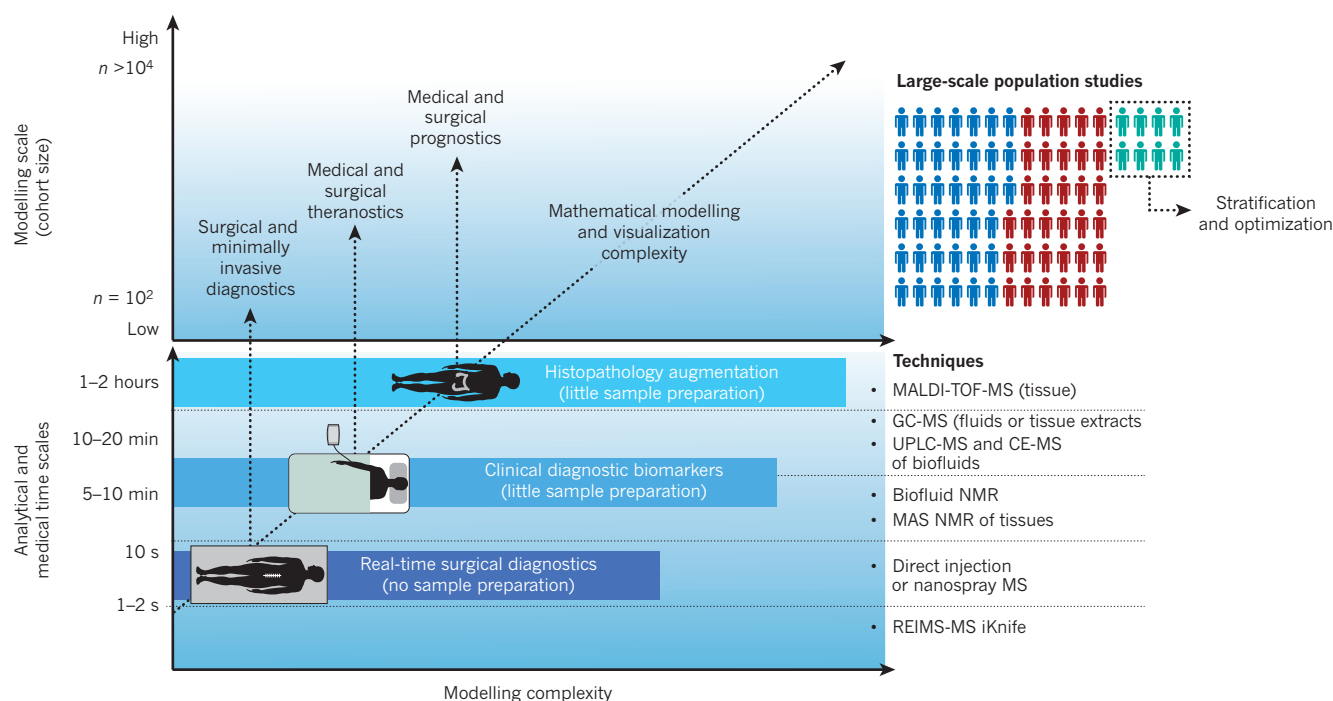


Figure 1 | Technology platforms and analytical timescales for patient journey phenotyping, diagnostic and prognostic biomarker discovery, and population disease-risk biomarker modelling. Different analytical technologies (Box 1) can be applied to a variety of clinically derived biosamples, and the choice of technology is dependent on the timescale for reaching a solution to the clinical problem, as well as the analytical performance characteristics of the technology. Thus, surgical problems require either real-time or near real-time solutions for clinical decision-making, whereas histopathological augmentation has a multi-hour timescale. The cohort size for predictive modelling for technique optimization depends on the biological information obtained. Epidemiological problems, such as

public health-care epidemiology and identifying new biomarkers as well as surgical risk stratification and pre-operative optimization involve the analysis of hundreds or thousands of samples from different populations, the transfer of information can cause bottlenecks to data processing and total-cohort analysis. In the critical care or surgical setting, large-scale population studies can also be used to identify populations at risk of surgical morbidity or a poor outcome. GC, gas chromatography; iKnife, intelligent knife; MALDI-TOF-MS, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry; MAS-NMR, magic-angle-spinning-nuclear-magnetic-resonance; MS, mass spectrometry; REIMS, rapid evaporative ionization mass spectrometry; UPLC, ultraperformance liquid chromatography.

drugs and gut-microbial activities⁸. The local phenotypic expression of the network interactions is obtained by analysing samples in these compartments, such as plasma or urine — which are the two most widely used clinical diagnostic fluids (Fig. 2) — or tissue. This analysis leads to the generation of a series of static snapshots of metabolic activity that can be difficult to interpret in isolation, unless there is overt metabolic disease, because of the background presence of physiological variability. Obtaining time series of samples from individuals who are undergoing diagnostic or prognostic evaluation, or at different stages of a disease process allows a longitudinal metabolic pattern or trajectory to emerge that carries much more information on site, severity and — potentially — mechanism of damage⁹. Similar arguments can be applied to responses to therapy. However, in reality, only a relatively small number of tissue or fluid types can be sampled, and the exhaustive analysis of these samples by advanced metabolic and spectroscopic techniques (Box 1) still gives only ‘islands of information’ that represent local activities (tissue or specialized biofluids) (Fig. 2) or systemic activities that affect the extracellular environment (urine and plasma). Thus, one of the challenges of metabolism-based ‘top-down’ systems biology¹⁰ is to try to build mathematical bridges between these islands to create system-level models that, in turn, can be used to generate biochemical or medical hypotheses for further testing using ‘bottom-up’ systems-biology methods. Furthermore, urine and plasma samples carry very different information sets on various molecules and pathways, representing numerous systemic timescales. For example, data from plasma provide a description of the metabolic system at the time of sampling, although persistent alterations induced by dietary or chronic interventions may also be detected. By contrast, information from urine is time-averaged because of its collection and storage in the bladder. There are also multiple complex physicochemical interactions and differences

in analytical matrix properties that not only determine how samples need to be prepared and analysed, but also carry other types of dynamic diagnostic information that is not demonstrated by simple (molecular identity and concentration) compositional analysis. Uniquely, NMR spectroscopic approaches do not disturb these complex perturbations in dynamic physicochemical interactions between molecules in biofluids, whereas this occurs of necessity in mass spectrometry. However, from a diagnostic point of view, dynamic chemical features have received relatively little attention in comparison with purely compositional biomarker analysis, despite the fact that some biofluids (such as semen) are highly reactive post collection due to intrinsic enzymatic activities¹¹. In terms of biological data generation, metabolomic and related methods are highly complementary to other ‘omics’ tools such as genomics, metagenomics, proteomics and transcriptomics, each of which covers different aspects of systemic and cellular-activity space, and all of which are interrelated. Systems-biology approaches also seek to integrate these data sets by using appropriate multivariate statistical and network modelling to obtain a more holistic view of human disease, although in the clinical environment multiple omics screening has, to date, rarely been feasible. No one tool, metric or platform gives a complete biological picture of a condition, and all approaches generate hypotheses that need rigorous testing and validation in the field. An advantage to metabolic profiling is that the relatively low cost per assay or procedure means it lends itself to large-scale testing and, in a clinical setting, useful information from commonly available samples (such as urine and plasma) can be obtained.

Overview of clinical applications

Metabotyping approaches have been used widely in animal models of disease, drug toxicity and drug action, resulting in many advances in the

BOX 1

Technology

NMR spectroscopy and mass spectrometry are the main techniques that are used for the metabolic profiling of biofluids (for example, urine, blood plasma, amniotic fluid and cerebrospinal fluid) and tissues in the form of extracts or intact biopsies^{94,95}. New solid-state NMR methods allow analysis of samples of less than 1 mg, thereby allowing detailed studies of tissue heterogeneity, such as between tumour tissue and tumour margins⁹⁶.

Both NMR and mass spectrometry can simultaneously identify and quantify information on a wide range of small molecules with good analytical precision and accuracy, and require only a small amount of sample (typically 10–400 µl). NMR spectroscopy is highly reproducible, with a detection limit in the sub-micromolar range. All hydrogen-containing metabolites in a biofluid are detected simultaneously and non-destructively with little sample preparation. Mass spectrometry has much lower detection limits, but it is destructive and a more targeted approach is often needed with prior separation of metabolites, using either chromatography or capillary electrophoresis⁹⁷. Thus, mass-spectrometry approaches tend to be less reproducible, more platform-dependent and susceptible to variability. Some of the earliest clinical studies used gas-chromatography–mass-spectrometry, especially for detection of inborn errors of metabolism⁹⁸, and, although still widely used, the requirement for chemical derivation of the sample to allow metabolite volatilization imposes limitations on its widespread use in clinical diagnostics. Liquid chromatography, particularly ultra-high-performance liquid chromatography (UPLC), is being used increasingly for metabolic profiling.

Chemometrics — multivariate statistics applied to chemical data — are used in clinical metabonomics to reduce the dimensionality of complex spectroscopic data sets, and to identify biochemical patterns that relate to a disease or an intervention. Linear-projection methods, such as principal-components analysis and partial-least-squares discriminant analysis, are commonly used to map samples on the basis of their biochemical similarity and to extract patterns of metabolites that relate to a particular disease¹⁴. Principal-components analysis is used extensively in metabonomics. This technique transforms the data descriptors into a set of linear combinations of the original features based on decreasing levels of variance. Any clustering seen is based on the data alone, and there is no pre-assignment of sample classes. Alternatively, in ‘supervised’ methods, multiparametric data sets can be modelled so that the class of separate samples (a ‘validation’ set) can be predicted based on a series of mathematical models derived from the original data or ‘training’ set. Partial-least-squares analysis is a

widely used supervised method (using a training set of data with known end points). This method relates a data matrix containing independent variables from samples, such as spectral intensity values (an X matrix) to a matrix containing dependent variables (for example, measurements of response) for those samples (a Y matrix). Partial least squares can also be combined with discriminant analysis to establish the optimal position to place a surface that best separates classes. Other popular chemometric methods include hierarchical clustering, self-organizing maps and neural networks^{14,95}.

Statistical spectroscopy is a form of computational modelling that is used to enhance biomarker recovery, allowing improved information extraction from a set of spectra. This method generally operates on defining correlation structures between variables (signals) that are found to be discriminatory between sample classes. Highly correlated signals are likely to come from the same molecule or from molecules regulated by the same metabolic pathway. Statistical total correlation spectroscopy is used to identify correlated signals within a data set for biomarker identification. NMR and mass spectrometry possess high complementarity in molecular-structure elucidation studies. In many metabonomic studies, multiple samples with a wide range of biochemical variation are available for both NMR and mass-spectrometry analysis, creating an opportunity for statistical analysis of signal amplitude co-variation between the two sets of data. Statistical heterospectroscopy is an extension of statistical total correlation spectroscopy for the co-analysis of multispectroscopic data sets, which have been acquired from multiple samples. The statistical heterospectroscopy approach, originally developed for NMR and mass-spectrometry correlation, can be used if any two or more independent spectroscopic data sets from any source are available for any sample cohort⁹⁹.

A new and exciting approach that uses mass spectrometry is the analysis of smoke from a cauterization device used in surgery to identify the exact type of tissue being investigated^{92,93}. By using a combination of new sampling methods, high-speed mass spectrometry and chemometrics for classification purposes, it is possible to identify different types of tissue in real time during a surgical procedure a development known as the intelligent knife (i-knife). A parallel application of mass spectrometry is to locate molecules within a sample as an imaging technique by using ionization methods based on laser ablation from the tissue surface⁹⁹. In combination with chemometric-enhanced information recovery, this has led to the possibility of an augmented histological assessment¹⁰⁰.

development of modelling techniques, chemometrics and ways to identify new biomarkers^{12–15}. But as the technology and modelling platforms have matured and improved, there has been a shift towards the implementation of clinical studies. Hence, in this Review, we focus on outlining specific areas that have the potential to have significant impacts on translational medicine and clinical delivery in the hospital environment. These include screening patients with established diseases for the detection of new biomarkers to aid in clinical classification. We are now in a position to deliver a systems-biology framework for complex clinical problems that are often compromised by extreme gene–environment variation. Early studies that applied metabolic profiling to clinical conditions were largely focused on identifying biomarkers, and were typically hindered by small group sizes and technical constraints. However, these studies paved the way for the wider use of metabolotyping approaches to further the understanding of systemic disease, as well as diagnostic and prognostic biomarker discovery. Indeed, the general principles of metabolotyping approaches had already been well-demonstrated by the 1980s, particularly for overt

metabolic disease such as that seen in type 2 diabetes¹⁶ or inborn errors of metabolism (based on ¹H-NMR-spectroscopy-derived urine and serum profiles)¹⁷. It was also shown quite early on that biochemically relevant systemic information could be recovered, such as the increase in serum alanine and the reduction in branched-chain amino acids (corresponding to decreased amino-acid gluconeogenesis and increased ketogenesis) that followed insulin withdrawal in people with type 2 diabetes. In addition, through time modification of plasma lipid and lipoprotein profiles, therapeutic optimization could also be monitored in this way¹⁶. The field has expanded to encompass epidemiological and population-scale studies¹⁸, and to take into account some of the complications of diabetes — such as vascular lesions — that lead to premature death¹⁹. Detailed cross-species metabolic analysis has uncovered information on the potential mechanisms underlying type 2 diabetes that relate to nucleotide metabolism, and to modulation of *N*-methylnicotinamide — which is conserved across rats, mice and humans²⁰. In the case of type 1 diabetes, a metabolic dysregulation of lipid and amino-acid metabolism was found to precede

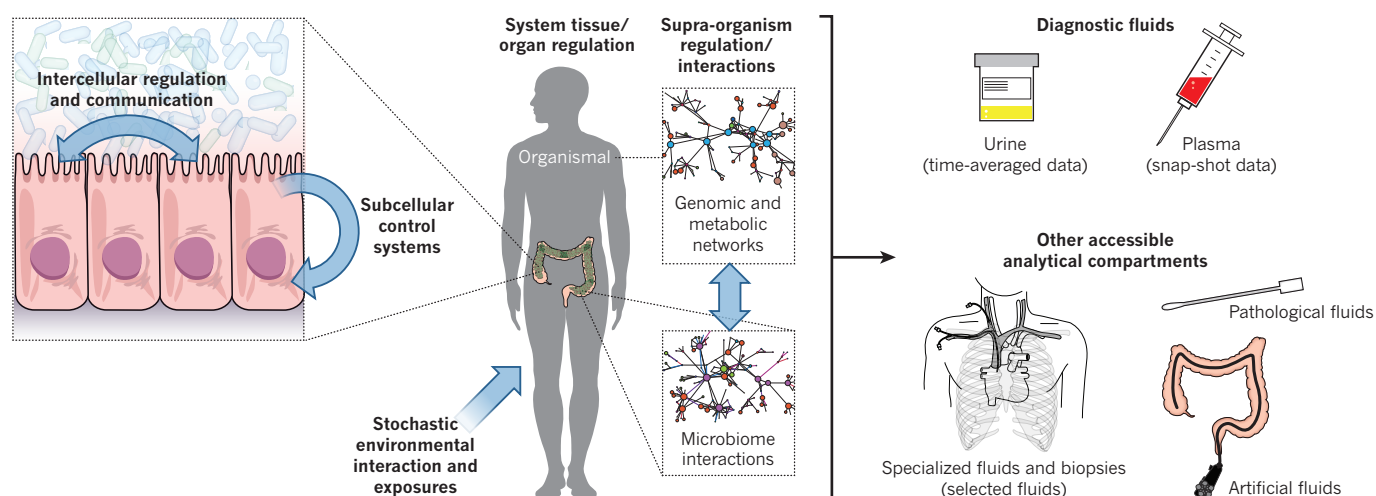


Figure 2 | Local and global metabolic interactions in relation to sampled compartments, fluids and their properties. In clinical settings, it is only possible to obtain limited ‘sampling islands’ for metabolic measurements. Within the body, there is a complex and dynamic continuum of metabolic interactions, from the subcellular level through multiple layers of biomolecular organization up to the whole supraorganism system, including the symbiotic microbiome components. At the system tissue and organ level, multicellular interactions occur through time and space through the secretion of biochemical products, as well as hormone and neurological control of function and physiological homeostatic regulation. Environmental and exogenous factors, including lifestyle, diet, drug therapy and the microbiota, all influence metabolism. For example, the microbiota, as part of the supraorganism, has a commensal and symbiotic relationship with tissues of the gut; the body’s interactions with pathogens and parasitic organisms, as well as quorum sensing, also have a role. At the intercellular level, signalling molecules and transporter systems coordinate functions and metabolic flux between cells. Finally, within the cell itself, enzymes require specific substrates and cofactors; biochemical conversions in organelles are topographically constrained, and the metabolome requires specific functional pathway units. The two most accessible components are urine and plasma, but they carry different system information sets as a result of these different regulation

and control systems. Because urine is stored in the bladder, it represents time-averaged data and has the following physiological characteristics: a variable pH, ionic strength and osmolarity; a high dielectric constant; an extreme dynamic concentration range (more than 10^{11}); thousands of molecules of less than 1 kDa; metal complexes and supramolecular aggregates; many small proteins; high enzyme activities in pathological states; and a dynamically reactive matrix. Plasma, however, provides snap-shot data and has the following characteristics: relatively constant pH, ionic strength and osmolarity, a lower bulk dielectric constant, a high dynamic concentration range (more than 10^5), hundreds of molecules both smaller and larger than 1 kDa; metal complexes and supramolecular complexes; a multi-compartment multi-diffusional matrix; and many large proteins and protein complexes. There are also a series of specialized secretory and pathological fluids that can be sampled and give, on spectroscopic analysis, more localized biochemical information specific to tissue injury. Specialized fluids are cerebrospinal, thyroid, saliva (sublingual, parotid and submaxillary), respiratory washings, gastric, bile, pancreatic, amniotic, follicular, milk, seminal vesicle, prostatic, epididymal and semen. Artificial fluids include bronchiolar lavage fluid, peritoneal dialysates, haemodialysates, faecal water, rectal dialysates, cell extracts and cell supernatants. Pathological fluids include ascites, pus, cystic fluid and effusions (malignant and infective).

onset of the disease²¹, whereas a panel of five branched-chain amino acids was found to be predictive of type 2 diabetes²². Unsurprisingly, insulin resistance, and both type 1 and type 2 diabetes, have been the subject of intensive metabolic investigation for many years, and the contributions of metabonomics and metabolomics have been reviewed extensively²³.

In the field of cancer (reviewed on page 364 of this issue), most studies were originally centred on extracts of the tumour tissue itself, and ¹H-NMR spectroscopy coupled to pattern recognition methods showed the ease with which discrete cancer-tissue types could be discriminated²⁴. Perhaps what is now more clinically relevant is the identification of potential cancer biomarkers in biofluids, including successful mapping of plasma ovarian cancer signatures — which are characterized by an altered pattern of ceramides and lysophospholipids, increased ketone bodies, and decreased alanine, valine and low-density lipoproteins^{25,26}. Patients with lung cancer have also been distinguished from a control group by their low urinary levels of hippurate and trigonelline, together with elevated D-3-hydroxyisovalerate, α-hydroxyisobutyrate and N-acetylglutamine²⁷. An inverse relationship between endometrial cancer and the metabolites stearic acid and serum acylcarnitines has been identified²⁸, and dysregulation of acylcarnitines also has a role in kidney cancer²⁹. Some metabonomic studies have found that models built on serum metabolite profiles perform better in terms of sensitivity than conventional markers, such as carcinoembryonic antigen in colorectal cancer³⁰ for predicting early-stage tumours (stage 0–2). Other cancer metabonomic studies — including those on ovarian and breast cancers^{25,31}, and renal-cell carcinoma³² — have shown promise in differentiating early- from late-stage tumours. Perhaps even more exciting than the diagnostic potential is the

ability of metabolic models to predict clinical outcomes for certain cancers. Micrometastases were predicted in a study of people with breast cancer, in which patients who went on to develop metastases were shown to have higher levels of plasma glucose, proline, lysine, phenylalanine and N-acetylcysteine and lower levels of lipids³³. Similarly, a recent study showed that — based on pretreatment serum samples for 500 women with metastatic breast cancer — time to progression, overall survival and treatment toxicity could be predicted from the serum levels of phenylalanine and glutamate (higher) and glucose (lower) for a subset of patients who were HER2 positive, although correlation between pre-treatment serum profile and outcome was not possible for the general trial population³⁴. As for diabetes, there has been an eruption of metabolic research on the processes of onset and progression of tumour development over the past decade, as well as identification of cancer biomarkers, and this has been comprehensively reviewed with respect to cellular biochemistry³⁵, therapeutic target discovery³⁶ and tumour typing^{32,37}. There have also been advances in metabolotyping as a tool for basic cardiac research³⁸ in the ability to predict cardiovascular events in baseline profiles of individuals at risk of coronary artery disease³⁹; as well as for understanding the origins of pathology, including the complex environmental and non-infectious microbiological triggers of disease⁴⁰. For the most part, diagnostic methods have been developed for serum and urine. However, for certain classes of disease — such as lung disease — gas chromatography methods for characterizing volatile components of exhaled breath condensate have shown considerable promise. For example, children with asthma and allergic rhinitis were distinguished from controls based on the alkane and aldehyde composition of breath condensate⁴¹; and differentiation

between each of the stages from 1 and 3 in chronic obstructive pulmonary disease identified ketones, methyl-branched alkanes and alcohols, in exhaled breath among other compounds⁴². An advantage of metabolic profiling is that multiple compartments and fluids can be analysed to give complementary information on systemic dysfunction. Thus, the metabolic signature of lung disease also extends to urine and serum. Urinary concentrations of the tricarboxylic-acid (TCA) cycle intermediates α -ketoglutarate, succinate, fumarate and *cis*-aconitate were found to be differential between people with stable and unstable asthma⁴³, whereas decreased serum lipoproteins and *N*-dimethylglycine, and increased glutamine, 3-methylhistidine and branched-chain amino acids have been associated with chronic obstructive pulmonary disease⁴⁴.

Human metabolic phenotypes, and multiple disease processes, are highly dependent on gut-microbial activity. An emerging area for metabolic profiling is the characterization of the functional properties of the gut microbiome. This is the combined genomic composition (more than 3.3 million genes⁴⁵) of several thousand species that make up the gut microbiota, and varies with age and between human populations⁴⁶. Abnormalities of the gut microbiome have been associated with a remarkable variety of human conditions, ranging from obesity and diabetes to autoimmune diseases and neuropsychiatric disorders⁴⁷. Metabotyping of inflammatory bowel diseases has been carried out on urine, plasma and faecal samples both to characterize the metabolic consequences of ulcerative colitis and Crohn's disease and to identify disease-induced changes in the metabolites deriving from the gut microbiota^{48–50}. Differences in the levels of metabolites, such as the short-chain fatty acids, 4-cresyl sulphate and hippurate are indicative of a perturbed microbiota in inflammatory bowel conditions, whereas altered levels of TCA-cycle intermediates and amino acids reflect a shift in energy balance. Crucially, the activities of the gut microbiota influence the host metabolic phenotypes⁵¹ through a series of complex signalling axes that connect to multiple host compartments, including the liver and brain⁴⁷, as well as the immune system⁵². The metabolic axes involve bile-acid (which are themselves heavily metabolized by the microbiota) dependent signalling, binding to a variety of nuclear receptors, such as those in the liver, which in turn affect host gene-expression profiles⁵³. There are also signalling axes that involve gut-microbe-generated short-chain fatty acids (from colonic fermentation of polysaccharides and oligosaccharides), aromatic amines and acids (from aromatic amino-acid and protein putrefaction in the distal colon⁵⁴), as well as gut-microbe-derived links (through the endocannabinoid system) which affect host adipogenesis⁵⁵, and ultimately multiple CNS signalling axis connections. Children who are diagnosed with the neurobehavioural disorder autism manifest different urinary metabolite phenotypes compared with controls, including increased excretion of gut-microbial metabolites (such as phenylacetylglutamine and 4-cresyl sulphate) as well as altered amino-acid and nicotinic-acid profiles^{56,57}. The microbiome influences metabolism from birth, and early events in the development of the microbiome–host signalling axes can leave a lasting metabolic imprint. Babies that are born before 37 weeks gestation have a higher risk of developing metabolic syndrome and end-stage renal failure than those born at full term. Individuals who are born preterm can still be differentiated from those born at full term when they reach adulthood by profiling of the microbial degradation products choline, bile acids and acetylated glycoproteins⁵⁸. These metabolic signalling axes may form the basis of drug discovery and other therapeutic strategies designed to operate either directly on the microbiome⁵⁹ or indirectly through interactions with host metabolic pathways and immune signalling⁶⁰. Indeed, the microbiome may offer more druggable targets than the human genome, but determining this lies in future research and will be highly dependent on the successful application of metabotyping approaches to help elucidate these complex microbial–host symbiotic interactions.

Phenotyping patient journeys

All patients who enter the diagnostic environment undergo a series of tests that is designed to characterize and stage disease, as well as to select suitable therapies, which then result in either a 'good' or 'bad' clinical

outcome for the individual — a process known as the patient journey. By use of the advanced metabotyping methods already described, coupled to classic clinical diagnostic criteria, it is now possible to conceptualize a phenotypically enhanced patient journey in which multiple technology platforms are deployed throughout the patient-handling pipeline⁶¹. The first level of deployment is to create enhanced diagnostic biomarker profiles at each stage of the journey to assess how the patient responds to therapies, as well as to form differential diagnoses. However, by using pharmacometabonomic approaches⁶², it is also possible to consider the sum of many patient journeys, and to engage in prospective or prognostic analysis of patient outcomes. In pharmacometabonomic studies, pre-intervention profiles of biofluids, such as urine or plasma, are used to create mathematical models of therapeutic interventions (using cross-validated models (Box 1)) so that prognostic outcomes can be judged. Such studies have predicted xenobiotic hepatotoxicity in experimental animal models through pre-intervention urinary profiling⁶², and predicted drug (paracetamol) metabolism in humans using NMR-based spectroscopic profiling of urine (also demonstrating a complex connection between gut microbes and drug metabolic fate)⁶³. An example of a successful pharmacometabonomic patient-stratification model is the prediction of response to capecitabine therapy in patients with colorectal cancer, whereby high levels of serum polyunsaturated fatty acids were predictive of drug toxicity⁶⁴. The approach is also relevant to understanding surgical interventional outcomes (discussed later)⁶¹, and is generally well-suited to modelling multiple longitudinal congruent patient journeys and for the abstraction of patient stratification information to help inform decision-making.

In particular, patient-journey phenotyping lends itself to scalable and translatable models that can be applied to any acute hospital admission, covering a variety of disease states (Fig. 3). Furthermore, this generalized patient-journey phenotyping protocol can be applied in a drug-development testing environment in which we envisage the development of a phenotypically augmented clinical trial. Standard clinical-trial information would be supplemented by molecular data, leading to a significant enhancement of the mechanistic background to observed responder or non-responder phenotypes that are seen in many clinical trials. Nowhere in the clinic is the ability to deliver a rapid prognostic metric of clinical condition more important than in the critical-care setting. Here, a gain in minutes or hours when choosing and implementing a therapeutic strategy can mean the difference between life and death, and poor decisions carry substantial financial cost. Recent studies suggest that metabolic profiling tools can augment the identification of sepsis based on a set of serum acylcarnitines and glycerophosphatidylcholines measured by liquid-chromatography–mass-spectrometry⁶⁵, whereas serum levels of triacylglycerides, glucose and glutamate are predictive of survival in patients who experienced trauma⁶⁶.

Mapping metabolic phenotypes during surgery

Surgery is, by definition, the most personalized of all therapeutic options delivered in current clinical practice, yet there is a lack of molecular diagnostic and prognostic instruments that are required for modern precision-based surgical practice¹. As a result, clinical decisions are made on the basis of calculations of surgical risk using single or univariate biomarkers or data from retrospective logistic regression models. Surgery also poses other challenges. First, clinical diagnostic phenotypes in surgery are often hard to define precisely and tend to be inadequate for omics-based research. The best example of this is found in the management of patients who have experienced major trauma, in whom injuries are highly heterogeneous and exert a variable and often unpredictable inter-individual effect⁶⁷. Secondly, patients who undergo modern surgery are exposed to a large and variable environmental load of operative drugs and bacteria, as well as nutritional optimization strategies, which cannot be objectively measured by standard biochemical assays⁶⁸. While surgical outcomes have improved, the population has aged, and patients increasingly have a more diverse range of co-morbidities, with higher associated rates of malnutrition⁶⁹, polypharmacy⁷⁰ and medical interventions.

Finally, the operating environment presents a unique and stringent set of requirements for molecular platforms. Not only must they be able to objectively quantify the large number of rapidly changing environmental influences on patient outcome during a single operative journey, but they must be rapid, highly reliable, inexpensive and physically able to function under the conditions of an operation. Longitudinal metabolotyping describes a metabonomic expression of an individual's metabolism, the trajectory of which can be accurately measured as a patient passes through a multivariate operative journey⁶¹. These data can be integrated into current clinical data sets to augment surgical decision-making, and deviations from a personalized trajectory can be used to detect early clinical deterioration, predict risk or guide intervention. There is now clinical evidence that this approach has merit. Mass-spectrometry-targeted analysis of 69 serum metabolites in a large population of patients (478 patients) who were undergoing cardiac surgery was able to accurately predict a poor operative outcome over a mean follow-up period of 4.3 ± 2.4 years⁷¹. Short-chain dicarboxylacylcarnitines, ketone-related metabolites and short-chain acylcarnitines were all independently predictive of an adverse outcome after multivariate adjustment. However, personalized or stratified approaches to health care must not only focus on mammalian biology. Initial work suggests that a metabonomic strategy is also able to predict early-onset systemic inflammatory response syndrome (SIRS) in those patients who are exposed to major trauma. Partial-least-squares discriminant analysis was also able to clearly distinguish between patients

with SIRS and multi-organ dysfunction syndrome, according to variation in levels of carbohydrate, amino acids, glucose, lactate, glutamine signals, fatty acyl chains and lipids⁷². Finally, metabonomics has had a particular impact on experimental models of kidney transplant surgery, in which it is being assessed for its ability to predict graft failure⁷³, and end-organ drug toxicity⁷⁴ and to assess hypoxic injury in cadaveric specimens. It has also been widely used in experimental models of liver⁷⁵ and gut⁷⁶ transplantation. Metabonomics may have a significant impact on transplant surgery, in which there is a demand for rapid molecular diagnostics within the operating room to predict graft suitability and survival.

The British surgeon Joseph Lister first described the importance of antiseptics more than 100 years ago, yet the wider role of gut bacteria in surgical health is only just being recognized and little is known about the importance of commensal bacteria to post-operative recovery. This is largely because clinicians rely on culture-dependent analysis to glean useful information about pathogenic organisms, or complex and dynamic ecosystems. Rapid, culture-independent analytical approaches for the detection of species-specific changes in pathogenic or commensal bacteria are therefore of particular use in surgery. However, clinicians need to know not only which bacteria are present, but also what they are doing. Global metabolotyping therefore extends beyond the model of culture-independent analysis — by allowing the exploration of the functional and symbiotic biochemical relationship between humans and microbiota — as patients progress through their recovery phase. For example, surgical

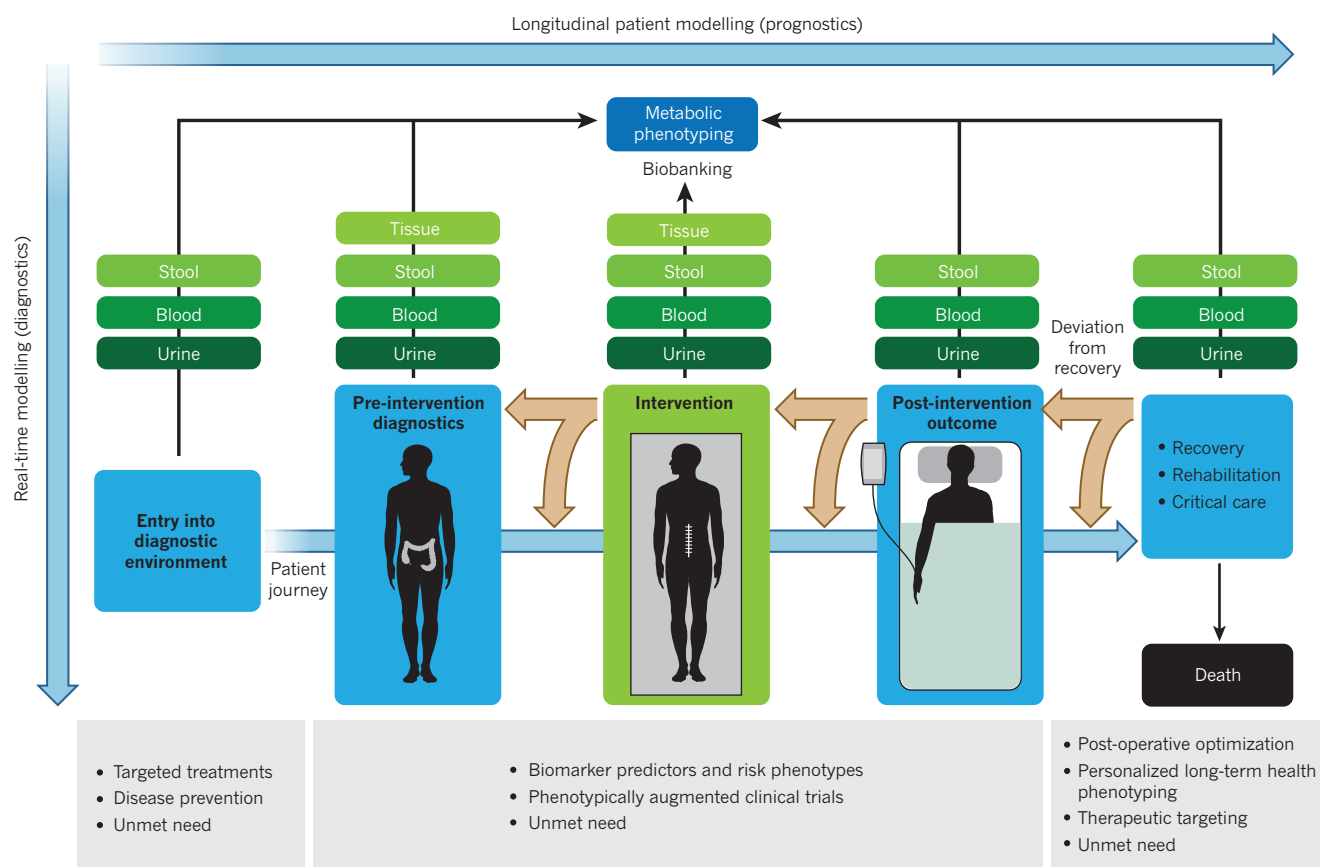


Figure 3 | Phenotyping the patient journey and phenotypically augmented clinical trials. Patients enter the diagnostic environment either through community admission, electively, as an acute case or as an emergency. At any point in the patient journey, there are multiple opportunities for metabolic phenotyping using technologies such as mass spectrometry or NMR spectrometry. Sections of these samples can also be stored in biobanks prior to analysis for use in future research. Deviation from recovery can occur at any point in the patient journey, and samples can be taken again. This analysis can be used to enhance differential diagnosis, therapeutic responses and long-term outcomes of therapy. Taking biosamples also provides real-time diagnosis and

prognosis to enhance clinical decision-making⁶¹. Before admission to hospital, this can be used for disease prevention and after intervention to optimize recovery. However, current biomarkers have left several areas of unmet clinical need in personalized prevention and therapeutic strategies, and in the delivery of sensitive and specific diagnostics and prognostic platforms for both surgical and medical diseases. By modelling congruent longitudinal journeys using pharmacometabonomic approaches⁶² it is possible to derive prognostic biomarker predictors and risk phenotypes that allow patient stratification, or that can give mechanistic information relating to therapeutic responder or non-responder status.

bypass of the foregut is increasingly used to treat obesity and its metabolic complications, such as type 2 diabetes⁷⁷. There are now convincing animal and human data to suggest that this surgical bypass fundamentally disrupts the distal gut microbiome, and that the subsequent disruption to the metabolome can be objectively measured using both NMR and mass-spectrometry-based approaches^{78,79}. Following bariatric surgery, there is a marked dysregulation in the gut microbiota from a landscape dominated by Firmicutes and Bacteroidetes to one dominated by proteobacteria^{78,80}. Metabolic profiles reflect this shift in the community with a persistent alteration in urinary, serum and faecal levels of cresols, indoles and biogenic amines^{78,79}. This suggests that personalized biomarkers for the prediction of long-term weight loss will have to account for the gut microbiome, and that the long-term health consequences of permanently altered enteric flow have yet to be fully defined⁷⁸. The gut microbiome is influenced by nearly all medical peri-operative therapeutic and risk-reduction strategies (for example, broad-spectrum antibiotic use), yet the effect of this on operative morbidity is not yet understood. Therefore, the clinical deployment of metabonomic technologies that provide real-time functional insight into the human–microbiome surgical health axis during surgery will have a significant effect. It is likely that this will be crucial to older people, who are the most vulnerable — as well as the fastest growing — group of patients who have surgery. It is well-established that stable age-related changes occur in gut-microbiome function⁸¹. Recent data suggest the presence of an age-related diet–microbiome–health axis, in which the gut microbiome is markedly different in patients living in residential care than those living in the community⁸². This could be linked to dietary changes, as well as to objective measures of frailty and poor health. The same is probably true of patients who are subjected to long stays in hospital, suggesting that measures of gut health are urgently required, and that new models of pre-operative nutritional optimization are needed that include metabonomic measures of gut health. Moreover, surgical excision of the colon may have much wider metabolic consequences for older people than currently understood.

Equally, there is a need for real-time interventional precision biomarkers that improve the quality and efficacy of the surgical excision itself. Oncological surgery is still based on Halsted's principles of oncological clearance, and the technique has remained largely unchanged since it was introduced. Clearance is typically based on arbitrary measurements that are not objective and are often inadequate, with potentially serious implications for the patient. For example, more than 20% of breast cancers require re-excision for positive margins⁸³. Metabonomics permits near real-time analysis with minimal sample preparation on very low volume samples, and several analytical platforms offer a potential solution to this problem. Magic-angle-spinning (MAS) ¹H-NMR spectroscopy is able to robustly determine the difference between benign and malignant tissue from patients with breast and colon cancer with a high degree of sensitivity and specificity⁸⁴. This has been extensively used in the analysis of brain tumours to differentiate between malignant tumour types⁸⁵, and these detailed biochemical profiles can then be related to the lower-resolution spectra obtained *in vivo* using NMR spectroscopy⁸⁶. This significantly improves MRI-based characterization of grade IV glioblastomas, metastases, medulloblastomas, lymphomas and glial tumours. Low concentrations of citrate and high concentrations of choline-containing compounds are metabolic characteristics that have been observed by NMR spectroscopy of prostate-cancer tissue. A similar approach has, therefore, been used in prostate cancer, in which this technique demonstrated an overall accuracy of between 93% and 97% for detecting the presence of prostate-cancer lesions⁸⁷.

In a study⁸⁸ to find regulatory genes with the potential for targeted therapies, the gene products acetylcytrate lyase and m-aconitase were both found to be predictive of significantly reduced citrate level. In the same study, which used 133 fresh-frozen samples from 41 patients undergoing radical prostatectomy, the two genes whose expression most closely accompanied the increase in choline-containing compounds were *PLA2G7* and *CHKA*. Thus, MAS-NMR spectroscopy, when incorporated into a systems-level analysis, provides new insight into cancer disease

mechanisms, and it is conceivable that MAS-NMR spectroscopy can be performed within a 10–20-minute time frame if facilities are co-located near to, or within, clinical environments. Therefore, this analysis has translational potential as a clinical resource for rapid diagnostics within either the outpatient clinic or the operating room environment. However, few technologies exist that are significantly faster than frozen-section histopathological analysis for the intra-operative assessment of tumour margins. Mass spectrometry has been used to characterize intact biological tissues for more than 30 years, but the field gained real momentum in the late 1990s with the advent of matrix-assisted laser desorption/ionization (MALDI) imaging analysis of histological tissue sections. Mass-spectrometry imaging studies, including MALDI, have revealed the molecular fingerprint of tissues with metabolic constituents, that lipids and proteins have a high histological specificity, and identified a number of prognostic markers (both single and complex). Mass spectrometry imaging has been suggested as an alternative to frozen-section histology; however, the time demand of this type of analysis is currently several hours per sample, even at coarse (100 µm) resolution⁸⁹. Multivariate, chemically augmented histology of this type therefore has two significant benefits over current histological staining methodologies. First, it provides instantaneous tissue identification, which allows interactive and feedback-controlled surgical and diagnostic interventions. Second, there is no inter-operator reproducibility of histological data, which can be exacerbated by low quality (for example, smeared) histological sections.

In contrast to mass-spectrometry imaging, rapid evaporative ionization mass spectrometry (REIMS) was developed exclusively for *in situ* analysis — even for the *in vivo* chemical characterization of tissues⁹⁰. REIMS was developed as a result of the discovery that all surgical instruments that use thermal evaporation approaches (including electrosurgery, laser surgery, radiofrequency ablation and microwave ablation) ionize the molecular constituents of biological tissues. The subsequent combination of surgical instruments with mass spectrometry has yielded an approach capable of identifying tissues and their pathological subtypes during surgical or diagnostic interventions⁹⁰. Data generated from REIMS is strikingly similar to other imaging mass spectrometry (MALDI or desorption electrospray ionization) data⁹¹. Although REIMS is a more suitable technology for the surgical environment than MALDI imaging, the latter guarantees a high histological specificity with quantitative, and potentially automated, histopathological analysis of tissue specimens. Recent results suggest that the REIMS technology can be successfully implemented in the surgical environment⁹⁰. Although the spatial resolution of REIMS is limited by the hand-held nature of the probe and its geometry, 100 µm is generally achievable with rapid (less than 0.9 s) feedback to the operator. Initial analysis of adenocarcinoma of the gastrointestinal system and lungs together with hepatic metastases, primary tumours of the liver, pre-cancerous degenerations of colon mucosa and sentinel lymph-node mapping has shown more than 95% concordance with classical histology and a less than 1% false-negative rate⁹². The technology has also been successfully trialled in neurosurgical brain-tumour excision of astrocytomas, meningiomas and metastatic brain tumours, as well as healthy brain tissue, with similar sensitivity and specificity⁹³. Thus, mass-spectrometry metabonomics provides real-time, descriptive *in vivo* data that are directly comparable with post-interventional histological analysis.

Concluding remarks

Systems-biology approaches have allowed a deeper understanding of the metabolic and physiological function of the human symbiotic 'supraorganism' (which includes the sum total of the eukaryotic and prokaryotic genomes required for human health). Disorders of supraorganism function underlie the aetiology of many modern non-communicable diseases, and metabolism-based mechanistic understanding of these processes therefore has much to offer personalized health-care systems of the future. Super-system surgery is the influence of surgical interventions or trauma on this complex symbiotic network, and the resultant time-dependent disruption in microbial–mammalian co-metabolic pathways. To take full advantage of theories such as this,

new systems-medicine technologies for surgical biomarker and drug target discovery are required, and these are already being developed. Metabonomics is used to probe the real-world nature of biochemical functionality and is sensitive to both gene and environmental influences; it is, therefore, likely to be more practical than gene-based measurements of responses to therapy. Nevertheless, a multi-omics approach can provide more information than a single one. Thus, it is currently possible to integrate heterogeneous data sources (for example, metagenomic and metabonomic, transcriptomic or proteomic data sets) to provide a complete top-down overview of complex disease states. We affirm that metabonomics has the power to influence clinical decision-making in the hospital environment for both medical and surgical treatments. The exquisite sensitivity of metabolic profiles to different diseases and treatment options means that computer models can be generated to aid decision-making processes for the medical practitioner. Moreover, by recording a patient's metabolite as treatment progresses, it will be possible to monitor the beneficial or detrimental effects of treatment, so that, for example, drug regimes or diet can be altered and a prognosis of disease outcome can be made. However, future systems must be able to link omics-level data sets and clinical databases seamlessly, and incorporating electronic health records into experimental data sets would seem to be an essential, although formidable, task for the future. ■

1. Mirnezami, R., Nicholson, J. & Darzi, A. Preparing for precision medicine. *N. Engl. J. Med.* **366**, 489–491 (2012).
2. Gavaghan, C. L., Holmes, E., Lenz, E., Wilson, I. D. & Nicholson, J. K. An NMR-based metabonomic approach to investigate the biochemical consequences of genetic strain differences: application to the C57BL/10J and Alpk:ApfCD mouse. *FEBS Lett.* **484**, 169–174 (2000).
3. Holmes, E. *et al.* Human metabolic phenotype diversity and its association with diet and blood pressure. *Nature* **453**, 396–400 (2008).
This study is the first example of the metabolome-wide association study concept in which disease risk factors (such as elevated blood pressure) were analysed in relation to exploratory (NMR) spectroscopic data.
4. Holmes, E., Wilson, I. D. & Nicholson, J. K. Metabolic phenotyping in health and disease. *Cell* **134**, 714–717 (2008).
5. Nicholson, J. K. & Lindon, J. C. Systems biology: metabonomics. *Nature* **455**, 1054–1056 (2008).
6. Fiehn, O. Metabolomics—the link between genotypes and phenotypes. *Plant Mol. Biol.* **48**, 155–171 (2002).
7. Nicholson, J. K., Lindon, J. C. & Holmes, E. 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica* **29**, 1181–1189 (1999).
This article describes and defines metabonomics as a tool for studying systemic metabolic changes due to disease, stresses, physiological stimulus or genetic modification.
8. Nicholson, J. K. & Wilson, I. D. Understanding 'global' systems biology: metabonomics and the continuum of metabolism. *Nature Rev. Drug Discov.* **2**, 668–676 (2003).
9. Holmes, E. *et al.* Nuclear magnetic resonance spectroscopy and pattern recognition analysis of the biochemical processes associated with the progression of and recovery from nephrotic lesions in the rat induced by mercury(II) chloride and 2-bromoethanamine. *Mol. Pharmacol.* **42**, 922–930 (1992).
This article reports the first use of metabolic profiling approaches to follow longitudinal changes in systemic metabolism.
10. Loscalzo, J., Kohane, I. & Barabasi, A. L. Human disease classification in the postgenomic era: a complex systems approach to human pathobiology. *Mol. Syst. Biol.* **3**, 124 (2007).
11. Tomlins, A. M. *et al.* High resolution ¹H NMR spectroscopic studies on dynamic biochemical processes in incubated human seminal fluid samples. *Biochim. Biophys. Acta* **1379**, 367–380 (1998).
12. Patterson, A. D. *et al.* Metabonomics reveals attenuation of the SLC6A20 kidney transporter in nonhuman primate and mouse models of type 2 diabetes mellitus. *J. Biol. Chem.* **286**, 19511–19522 (2011).
13. Robertson, D. G., Reilly, M. D. & Baker, J. D. Metabonomics in pharmaceutical discovery and development. *J. Proteome Res.* **6**, 526–539 (2007).
14. Trygg, J., Holmes, E. & Lundstedt, T. Chemometrics in metabonomics. *J. Proteome Res.* **6**, 469–479 (2007).
15. Nevodomskaia, E., Mayboroda, O. A. & Deelder, A. M. Cross-platform analysis of longitudinal data in metabolomics. *Mol. Biosyst.* **7**, 3214–3222 (2011).
16. Nicholson, J. K. *et al.* Proton-nuclear-magnetic-resonance studies of serum, plasma and urine from fasting normal and diabetic subjects. *Biochem. J.* **217**, 365–375 (1984).
17. Iles, R. A., Snodgrass, G. J., Chalmers, R. A. & Stacey, T. E. Rapid screening of metabolic diseases by proton NMR. *Lancet* **2**, 1221–1222 (1984).
This article provides an early example of the power of non-targeted phenotyping for use in classification of metabolic diseases and for exploring pathway abnormalities in genetic disease.
18. Suhre, K. *et al.* Metabolic footprint of diabetes: a multiplatform metabolomics study in an epidemiological setting. *PLoS ONE* **5**, e13953 (2010).
19. Makinen, V. P. *et al.* ¹H NMR metabonomics approach to the disease continuum of diabetic complications and premature death. *Mol. Syst. Biol.* **4**, 167 (2008).
20. Salek, R. M. *et al.* A metabolomic comparison of urinary changes in type 2 diabetes in mouse, rat, and human. *Physiol. Genomics* **29**, 99–108 (2007).
21. Oresic, M. *et al.* Dysregulation of lipid and amino acid metabolism precedes islet autoimmunity in children who later progress to type 1 diabetes. *J. Exp. Med.* **205**, 2975–2984 (2008).
22. Wang, T. J. *et al.* Metabolite profiles and the risk of developing diabetes. *Nature Med.* **17**, 448–453 (2011).
23. Friedrich, N. Metabolomics in diabetes research. *J. Endocrinol.* **215**, 29–42 (2012).
24. Howells, S. L., Maxwell, R. J., Griffiths, J. R. Classification of tumour ¹H NMR spectra by pattern recognition. *NMR Biomed.* **5**, 59–64 (1992).
25. Fan, L. *et al.* Identification of metabolic biomarkers to diagnose epithelial ovarian cancer using a UPLC/QTOF/MS platform. *Acta Oncol.* **51**, 473–479 (2012).
26. Garcia, E. *et al.* Diagnosis of early stage ovarian cancer by ¹H NMR metabonomics of serum explored by use of a microflow NMR probe. *J. Proteome Res.* **10**, 1765–1771 (2011).
27. Carrola, J. *et al.* Metabolic signatures of lung cancer in biofluids: NMR-based metabonomics of urine. *J. Proteome Res.* **10**, 221–230 (2011).
28. Gaudet, M. M. *et al.* Analysis of serum metabolic profiles in women with endometrial cancer and controls in a population-based case-control study. *J. Clin. Endocrinol. Metab.* **97**, 3216–3223 (2012).
29. Ganti, S. *et al.* Urinary acylcarnitines are altered in human kidney cancer. *Int. J. Cancer* **130**, 2791–2800 (2012).
30. Nishiumi, S. *et al.* A novel serum metabolomics-based diagnostic approach for colorectal cancer. *PLoS ONE* **7**, e40459 (2012).
31. Slupsky, C. M. *et al.* Urine metabolite analysis offers potential early diagnosis of ovarian and breast cancers. *Clin. Cancer Res.* **16**, 5835–5841 (2010).
32. Lin, L. *et al.* LC-MS based serum metabonomic analysis for renal cell carcinoma diagnosis, staging, and biomarker discovery. *J. Proteome Res.* **10**, 1396–1405 (2011).
33. Oakman, C. *et al.* Identification of a serum-detectable metabolomic fingerprint potentially correlated with the presence of micrometastatic disease in early breast cancer patients at varying risks of disease relapse by traditional prognostic methods. *Ann. Oncol.* **22**, 1295–1301 (2011).
34. Tenori, L. *et al.* Exploration of serum metabolomic profiles and outcomes in women with metastatic breast cancer: a pilot study. *Mol. Oncol.* **6**, 437–444 (2012).
35. Griffin, J. L. & Shockcor, J. P. Metabolic profiles of cancer cells. *Nature Rev. Cancer* **4**, 551–561 (2004).
36. Tennant, D. A., Durán, R. V. & Gottlieb, E. Targeting metabolic transformation for cancer therapy. *Nature Rev. Cancer* **10**, 267–277 (2010).
This is an important study on the use of targeted metabolic analysis for understanding fundamental metabolic processes in cancer cells for the discovery of drug targets and strategies.
37. Spratlin, J. L., Serkova, N. J. & Eckhardt, S. G. Clinical applications of metabolomics in oncology: a review. *Clin. Cancer Res.* **15**, 431–440 (2009).
38. Griffin, J. L., Atherton, H., Shockcor, J. P. & Atzori, L. Metabolomics as a tool for cardiac research. *Nature Rev. Cardiol.* **8**, 630–643 (2011).
39. Shah, S. H. *et al.* Baseline metabolomic profiles predict cardiovascular events in patients at risk for coronary artery disease. *Am. Heart J.* **163**, 844–850 (2012).
40. Wang, Z. *et al.* Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* **472**, 57–63 (2011).
This article reports the major discovery of the potential involvement of gut-microbial metabolism in developing cardiovascular disease.
41. Caldeira, M. *et al.* Profiling allergic asthma volatile metabolic patterns using a headspace-solid phase microextraction/gas chromatography based methodology. *J. Chromatogr. A* **1218**, 3771–3780 (2011).
42. Fens, N. *et al.* Exhaled air molecular profiling in relation to inflammatory subtype and activity in COPD. *Eur. Respir. J.* **38**, 1301–1309 (2009).
43. Saude, E. J. *et al.* Metabolomic profiling of asthma: diagnostic utility of urine nuclear magnetic resonance spectroscopy. *J. Allergy Clin. Immunol.* **127**, 757–764 (2011).
44. Ubhi, B. K. *et al.* Metabolic profiling detects biomarkers of protein degradation in COPD patients. *Eur. Respir. J.* **40**, 345–355 (2012).
45. Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59–65 (2010).
46. Yatsunenkov, T. *et al.* Human gut microbiome viewed across age and geography. *Nature* **486**, 222–227 (2012).
47. Nicholson, J. K. *et al.* Host-gut microbiota metabolic interactions. *Science* **336**, 1262–1267 (2012).
48. Ooi, M. *et al.* GC/MS-based profiling of amino acids and TCA cycle-related molecules in ulcerative colitis. *Inflamm. Res.* **60**, 831–840 (2011).
49. Williams, H. R. *et al.* Characterization of inflammatory bowel disease with urinary metabolic profiling. *Am. J. Gastroenterol.* **104**, 1435–1444 (2009).
50. Marchesi, J. R. *et al.* Rapid and non-invasive metabonomic characterization of inflammatory bowel disease. *J. Proteome Res.* **6**, 546–551 (2007).
51. Li, M. *et al.* Symbiotic gut microbes modulate human metabolic phenotypes. *Proc. Natl Acad. Sci. USA* **105**, 2117–2122 (2008).
This article reports the first demonstration of statistical cross-omics integration to unravel gut-microbe-host metabolic interactions.
52. Hooper, L. V., Littman, D. R. & Macpherson, A. J. Interactions between the microbiota and the immune system. *Science* **336**, 1268–1273 (2012).
53. Swann, J. R. *et al.* Systemic gut microbial modulation of bile acid metabolism in host tissue compartments. *Proc. Natl Acad. Sci. USA* **108**, 4523–4530 (2011).

54. Holmes, E. *et al.* Therapeutic modulation of microbiota–host metabolic interactions. *Sci. Transl. Med.* **4**, 137rv6 (2012).
This article provides a comprehensive discussion of major gut-microbe–host metabolic interactions and possible therapeutic interventional strategies.
55. Muccioli, G. G. *et al.* The endocannabinoid system links gut microbiota to adipogenesis. *Mol. Syst. Biol.* **6**, 392 (2010).
56. Yap, I. K. *et al.* Urinary metabolic phenotyping differentiates children with autism from their unaffected siblings and age-matched controls. *J. Proteome Res.* **9**, 2996–3004 (2010).
57. Evans, C. *et al.* Altered amino acid excretion in children with autism. *Nutr. Neurosci.* **11**, 9–17 (2008).
58. Thomas, E. L. *et al.* Aberrant adiposity and ectopic lipid deposition characterize the adult phenotype of the preterm infant. *Pediatr. Res.* **70**, 507–512 (2011).
59. Gordon J. I. Honor thy gut symbionts: redux. *Science* **336**, 1251–1253 (2012).
This article provides an overview of the importance of the gut microbiome in the aetiopathogenesis of diverse non-infectious diseases.
60. Jia, W., Li, H., Zhao, L. & Nicholson, J. K. Gut microbiota: a potential new territory for drug targeting. *Nature Rev. Drug Discov.* **7**, 123–129 (2008).
61. Kinross, J. M., Holmes, E., Darzi, A. W. & Nicholson, J. K. Metabolic phenotyping for monitoring surgical patients. *Lancet* **377**, 1817–1819 (2011).
62. Clayton, T. A. *et al.* Pharmacometabonomic phenotyping and personalized drug treatment. *Nature* **440**, 1073–1077 (2006).
This report provides the first description of the use of pre-interventional metabolic profile models to predict interventional outcomes.
63. Clayton, T. A., Baker, D., Lindon, J. C., Everett, J. R. & Nicholson, J. K. Pharmacometabonomic identification of a significant host–microbiome metabolic interaction affecting human drug metabolism. *Proc. Natl Acad. Sci. USA* **106**, 14728–14733 (2009).
64. Backshall, A., Sharma, R., Clarke, S. J. & Keun, H. C. Pharmacometabonomic profiling as a predictor of toxicity in patients with inoperable colorectal cancer treated with capecitabine. *Clin. Cancer Res.* **17**, 3019–3028 (2011).
This study is the first example of pharmacometabonomic principles to predict drug toxicity in humans.
65. Schmerler, D. *et al.* Targeted metabolomics for discrimination of systemic inflammatory disorders in critically ill patients. *J. Lipid Res.* **53**, 1369–1375 (2012).
66. Cohen, M. J., Serkova, N. J., Wiener-Kronish, J., Pittet, J. F. & Niemann, C. U. ¹H-NMR-based metabolic signatures of clinical outcomes in trauma patients—beyond lactate and base deficit. *J. Trauma* **69**, 31–40 (2010).
67. Polinder, S., Haagsma, J. A., Toet, H. & van Beeck, E. F. Epidemiological burden of minor, major and fatal trauma in a national injury pyramid. *Br. J. Surg.* **99**, 114–121 (2012).
68. Alverdy, J. C., Laughlin, R. S. & Wu, L. Influence of the critically ill state on host–pathogen interactions within the intestine: gut-derived sepsis redefined. *Crit. Care Med.* **31**, 598–607 (2003).
69. Volkert, D., Saeglit, C., Gueldenzoph, H., Sieber, C. C. & Stehle, P. Undiagnosed malnutrition and nutrition-related problems in geriatric patients. *J. Nutr. Health Aging* **14**, 387–392 (2010).
70. Fitzgerald, S. P. & Bean, N. G. An analysis of the interactions between individual comorbidities and their treatments – implications for guidelines and polypharmacy. *J. Am. Med. Dir. Assoc.* **11**, 475–484 (2010).
71. Shah, A. A. *et al.* Metabolic profiles predict adverse events after coronary artery bypass grafting. *J. Thorac. Cardiovasc. Surg.* **143**, 873–878 (2012).
72. Mao, H. *et al.* Systemic metabolic changes of traumatic critically ill patients revealed by an NMR-based metabonomic approach. *J. Proteome Res.* **8**, 5423–5430 (2009).
73. Chen, J. *et al.* Metabonomics study of the acute graft rejection in rat renal transplantation using reversed-phase liquid chromatography and hydrophilic interaction chromatography coupled with mass spectrometry. *Mol. Biosyst.* **8**, 871–878 (2012).
74. Kim, C. D. *et al.* Metabonomic analysis of serum metabolites in kidney transplant recipients with cyclosporine A- or tacrolimus-based immunosuppression. *Transplantation* **90**, 748–756 (2010).
75. Legido-Quigley, C. *et al.* Bile UPLC-MS fingerprinting and bile acid fluxes during human liver transplantation. *Electrophoresis* **32**, 2063–2070 (2011).
76. Girlanda, R. *et al.* Metabonomics of human intestinal transplant rejection. *Am. J. Transplant.* <http://dx.doi.org/10.1111/j.1600-6143.2012.04183.x> (July 2012).
77. Fornari, F., Comis, V. R. & Lisboa, H. R. Bariatric surgery or medical therapy for obesity. *N. Engl. J. Med.* **367**, 474 (2012).
78. Li, J. V. *et al.* Metabolic surgery profoundly influences gut microbial–host metabolic cross-talk. *Gut* **60**, 1214–1223 (2011).
79. Mutch, D. M. *et al.* Metabolite profiling identifies candidate markers reflecting the clinical adaptations associated with Roux-en-Y gastric bypass surgery. *PLoS ONE* **4**, e7905 (2009).
80. Zhang, H. *et al.* Human gut microbiota in obesity and after gastric bypass. *Proc. Natl Acad. Sci. USA* **106**, 2365–2370 (2009).
81. Biagi, E., Candela, M., Fairweather-Tait, S., Franceschi, C. & Brigidi, P. Aging of the human metaorganism: the microbial counterpart. *Age (Dordr.)* **34**, 247–267 (2012).
82. Claesson, M. J. *et al.* Gut microbiota composition correlates with diet and health in the elderly. *Nature* **488**, 178–184 (2012).
83. Jeevan, R. *et al.* Reoperation rates after breast conserving surgery for breast cancer among women in England: retrospective study of hospital episode statistics. *Br. Med. J.* **345**, e4505 (2012).
84. Chan, E. C. *et al.* Metabolic profiling of human colorectal cancer using high-resolution magic angle spinning nuclear magnetic resonance (HR-MAS NMR) spectroscopy and gas chromatography mass spectrometry (GC/MS). *J. Proteome Res.* **8**, 352–361 (2009).
85. Opstad, K. S., Bell, B. A., Griffiths, J. R. & Howe, F. A. An investigation of human brain tumour lipids by high-resolution magic angle spinning ¹H MRS and histological analysis. *NMR Biomed.* **21**, 677–685 (2008).
86. Wright, A. J. *et al.* Ex-vivo HRMAS of adult brain tumours: metabolite quantification and assignment of tumour biomarkers. *Mol. Cancer* **9**, 66 (2010).
87. Wu, C. L. *et al.* Metabolomic imaging for human prostate cancer detection. *Sci. Transl. Med.* **2**, 16ra18 (2010).
88. Bertilsson, H. *et al.* Changes in gene transcription underlying the aberrant citrate and choline metabolism in human prostate cancer samples. *Clin. Cancer Res.* **18**, 3261–3269 (2012).
89. McDonnell, L. A. & Heeren, R. M. Imaging mass spectrometry. *Mass Spectrom. Rev.* **26**, 606–643 (2007).
90. Balog, J. *et al.* Identification of biological tissues by rapid evaporative ionization mass spectrometry. *Anal. Chem.* **82**, 7343–7350 (2010).
This article provides a description of the technology development and application of the ‘intelligent knife’ concept for real-time surgical diagnostics.
91. Guenther, S. *et al.* Electrospray post-ionization mass spectrometry of electrosurgical aerosols. *J. Am. Soc. Mass Spectrom.* **22**, 2082–2089 (2011).
92. Gerbig, S. *et al.* Analysis of colorectal adenocarcinoma tissue by desorption electrospray ionization mass spectrometric imaging. *Anal. Bioanal. Chem.* **403**, 2315–2325 (2012).
93. Schafer, K. C. *et al.* Real time analysis of brain tissue by direct combination of ultrasonic surgical aspiration and sonic spray mass spectrometry. *Anal. Chem.* **83**, 7729–7735 (2011).
94. Beckonert, O. *et al.* Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nature Protoc.* **2**, 2692–2703 (2007).
95. Lindon, J. C. & Nicholson, J. K. Spectroscopic and statistical techniques for information recovery in metabolomics and metabolomics. *Annu. Rev. Anal. Chem.* **1**, 45–69 (2008).
96. Wong, A. *et al.* Evaluation of high resolution magic-angle coil spinning NMR spectroscopy for metabolic profiling of nanoliter tissue biopsies. *Anal. Chem.* **84**, 3843–3848 (2012).
97. Jellum, E. *et al.* Application of glass capillary-column gas chromatography–mass spectrometry to the studies of human diseases. *J. Chromatogr.* **126**, 487–493 (1976).
98. Ramautar, R., Mayboroda, O. A., Somsen, G. W. & de Jong, G. J. CE-MS for metabolomics: developments and applications in the period 2008–2010. *Electrophoresis* **32**, 52–65 (2011).
99. Crockford, D. J. *et al.* Statistical heterospectroscopy, an approach to the integrated analysis of NMR and UPLC-MS data sets: application in metabonomic toxicology studies. *Anal. Chem.* **78**, 363–371 (2006).
100. Fonville, J. M. *et al.* Robust data processing and normalization strategy for MALDI mass spectrometric imaging. *Anal. Chem.* **84**, 1310–1319 (2012).

Acknowledgements The authors would like to acknowledge the National Institute of Health Research Biomedical Research Centre for funding clinical and surgical metabonomic projects in real-time diagnostics and chemical imaging at Imperial College London. We also wish to thank the MRC and NIHR for funding major programmes that relate to these studies, including the MRC-NIHR Phenome Centre (joint with Kings College London, Bruker Spectrospin and the Waters Corporation) and the Imperial NIHR Clinical Phenome Centre.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on this article at go.nature.com/jpozrc. Correspondence should be addressed to J.K.N. (j.nicholson@imperial.ac.uk).

Analyses of pig genomes provide insight into porcine demography and evolution

A list of authors and their affiliations appears at the end of the paper

For 10,000 years pigs and humans have shared a close and complex relationship. From domestication to modern breeding practices, humans have shaped the genomes of domestic pigs. Here we present the assembly and analysis of the genome sequence of a female domestic Duroc pig (*Sus scrofa*) and a comparison with the genomes of wild and domestic pigs from Europe and Asia. Wild pigs emerged in South East Asia and subsequently spread across Eurasia. Our results reveal a deep phylogenetic split between European and Asian wild boars ~1 million years ago, and a selective sweep analysis indicates selection on genes involved in RNA processing and regulation. Genes associated with immune response and olfaction exhibit fast evolution. Pigs have the largest repertoire of functional olfactory receptor genes, reflecting the importance of smell in this scavenging animal. The pig genome sequence provides an important resource for further improvements of this important livestock species, and our identification of many putative disease-causing variants extends the potential of the pig as a biomedical model.

The domestic pig (*Sus scrofa*) is a eutherian mammal and a member of the Cetartiodactyla order, a clade distinct from rodent and primates, that last shared a common ancestor with humans between 79 and 97 million years (Myr) ago^{1,2} (<http://www.timetree.net>). Molecular genetic evidence indicates that *Sus scrofa* emerged in South East Asia during the climatic fluctuations of the early Pliocene 5.3–3.5 Myr ago. Then, beginning ~10,000 years ago, pigs were domesticated in multiple locations across Eurasia³ (Frantz, L. A. F. *et al.*, manuscript submitted).

Here we provide a high-quality draft pig genome sequence developed under the auspices of the Swine Genome Sequencing Consortium^{4,5}, established using bacterial artificial chromosome (BAC)⁶ and whole-genome shotgun (WGS) sequences (see Methods and Supplementary Information). The assembly (Sscrofa10.2) comprises 2.60 gigabases (Gb) assigned to chromosomes with a further 212 megabases (Mb) in unplaced scaffolds (Table 1 and Supplementary Tables 1–3).

Genome annotation

A *de novo* repeat discovery and annotation strategy (Supplementary Fig. 8) revealed a total of 95 novel repeat families, including: 5 long interspersed elements (LINEs), 6 short interspersed elements (SINEs), 8 satellites and 76 long terminal repeats (LTRs). The relative content of repetitive elements (~40%, Supplementary Figs 9 and 10) is lower than reported for other mammalian genomes. The main repetitive element groups are the LINE1 and glutamic acid transfer RNA (tRNA^{Glu})-derived SINEs or PRE (porcine repetitive element). The expansion of PRE is specific to the porcine lineage. Phylogenetic analysis of LINE1 and PRE (Supplementary Figs 13 and 14) indicates that only a single lineage of each is currently active and that the main expansion of both LINE1 and PRE occurred in the first half of the Tertiary period. Smaller expansions, particularly in LINE1, have occurred since, but recent activity is very low (Supplementary Information).

Annotation of genes, transcripts and predictions of orthologues and paralogues was performed using the Ensembl analysis pipeline⁷ (Table 1 and Supplementary Figs 3–7). Further annotation for non-protein-coding RNAs (ncRNAs) was undertaken with another analysis pipeline (Supplementary Information and Supplementary Table 4).

Evolution of the porcine genome

Evolution of genes and gene families

To examine the mutation rate and type of protein-coding genes that show accelerated evolution in pigs, we identified ~9,000 as 1:1 orthologues within a group of six mammals (human, mouse, dog, horse, cow and pig). This orthologous gene set was used to identify proteins that show accelerated evolution in each of these six mammalian lineages (Supplementary Information). The observed number of synonymous substitutions per synonymous site (dS) for the pig lineage (0.160) is similar to that of the other mammals (0.138–0.201) except for the mouse (0.458), indicating similar evolutionary rates in pigs and other mammals. The observed dN/dS ratio (ratio of the rate of non-synonymous substitutions to the rate of synonymous substitutions) of 0.144 is between those of humans (0.163) and mice (0.116), indicating an intermediate level of purifying selection pressure in the pig. Genes showing increased dN/dS ratios in each lineage were analysed using DAVID⁸ to examine whether these rapidly evolving genes were enriched for specific biological processes. Most lineages show different fast-evolving pathways, but some pathways are shared (Fig. 1).

Immune genes are known to be actively evolving in mammals^{9,10}. Because many immune genes were not included in the analysis of 1:1 orthologues, we examined a randomly selected subset of 158 immunity-related pig proteins for evidence of accelerated evolution (Supplementary information). Twenty-seven of these genes (17%)

Table 1 | Assembly and annotation statistics

Assembly	Placed	Unplaced	Annotation*
Total length	2,596,639,456	211,869,922	21,640 protein-coding genes
Ungapped length	2,323,671,356	195,490,322	380 pseudogenes
Scaffolds	5,343	4,562	2,965 ncRNAs†
Contigs	73,524	168,358	197,675 gene exons
Scaffold N50	637,332	98,022	26,487 gene transcripts
Contig N50	80,720	2,423	

* Numbers refer to the annotation performed by Ensembl (release 67). Results of an independent annotation by the NCBI can be obtained from <http://www.ncbi.nlm.nih.gov/mapview/stats/BuildStats.cgi?taxid=9823&build=4&ver=1>.

† An improved ncRNA annotation with 3,601 ncRNAs and structured elements is available as a separate track in Ensembl version 70 and for download from <http://rth.dk/resources/rnannotator/susscr102>. N50, 50% of the genome is in fragments of this length or longer.

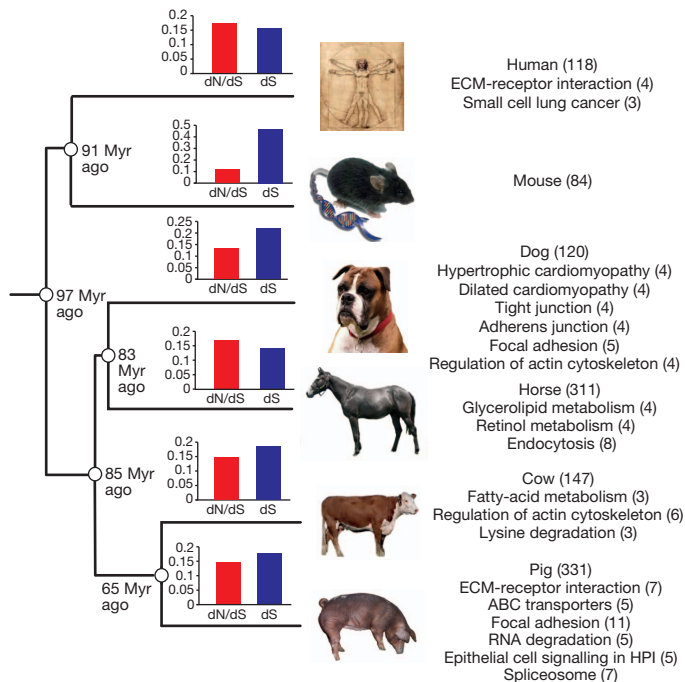


Figure 1 | Phylogeny of the six mammals used in the dN/dS analysis. KEGG pathways with genes that show accelerated evolution for each of the six mammals used in the dN/dS analysis. The bar charts show the individual dN/dS and dS values for each of the six mammals. The dN/dS and dS values refer to the time period of each of the six individual lineages. The number of proteins that show significantly accelerated dN/dS ratios in each lineage varies from 84 in the mouse to 311 in the pig lineage. Pathways significantly ($P < 0.05$) enriched within this group of genes are also shown with the number of genes shown in brackets. HPI, *Helicobacter pylori* infection.

demonstrated accelerated evolution (Supplementary Table 8). A parallel analysis of 143 human and 145 bovine orthologues revealed very similar rates of evolution (18% in human and 12% in cattle, respectively). Using a branch-site analysis, we detected accelerated evolution of amino acids in PRSS12, CD1D and TRAF3 specific to pig (positive selection on pig branch), as well as amino acids in TREM1, IL1B and SCARA5 specific to pig and cow (positive selection on the cetartiodactyl branch).

Further analysis of porcine immune genes (Supplementary Table 5) revealed evidence for specific gene duplications and gene-family expansions (Supplementary Tables 6 and 7). The analysis of this second cetartiodactyl genome indicates that some expansions are cetartiodactyl-specific (cathelicidin) whereas others are ruminant/bovine-specific (β -defensins, C-type lysozymes) or potentially porcine-specific (type I interferon, δ subfamily).

Pigs have at least 39 type I interferon (IFN) genes, which is twice the number identified in humans and significantly more than in mice. We also detected 16 pseudogenes in this family. Cattle have 51 type I IFNs (13 pseudogenes), indicating that both bovine and porcine type I IFN families have undergone expansion. This is particularly important for interferon subtypes δ (IFND), ω (IFNW) and τ (IFNT); pigs and cattle are evolving species-specific subtypes of IFND and IFNT, respectively. Both species are expanding the IFNW family and share many more IFNW isoforms than other species. Thus, expansion of interferon genes is not ruminant-specific as proposed earlier¹⁰, although duplication within some specific sub-families seems to be either bovine- or porcine-specific.

Within the immunity-related genes annotated, we found evidence for duplication of six immune-related genes: *IL1B*, *CD36*, *CD68*, *CD163*, *CRP* and *IFIT1*, and one non-immune gene, *RDH16*. The *CD36* gene is also duplicated in the bovine genome, whereas the *IL1B* gene duplication, where evidence for a partial duplication was

reported previously¹¹, is unique in mammals. Other key immune genes in the major histocompatibility complex, immunoglobulin, T-cell-receptor and natural killer cell receptor loci have been characterized in detail^{12–19} (Supplementary Information).

Another significant porcine genome expansion is the olfactory receptor gene family. We identified 1,301 porcine olfactory receptor genes and 343 partial olfactory receptor genes²⁰. The fraction of pseudogenes within these olfactory receptor sequences (14%) is the lowest observed in any species so far. This large number of functional olfactory receptor genes most probably reflects the strong reliance of pigs on their sense of smell while scavenging for food.

Conservation of synteny and evolutionary breakpoints

Alignment of the porcine genome against seven other mammalian genomes (Supplementary Information) identified homologous synteny blocks (HSBs). Using porcine HSBs and stringent filtering criteria, 192 pig-specific evolutionary breakpoint regions (EBRs) were located. The number of porcine EBRs (146, Supplementary Table 11 and Supplementary Fig. 16) is comparable to the number of bovine-lineage-specific EBRs (100) reported earlier using a slightly lower resolution (500 kilobases (kb)), indicating that both lineages evolved with an average rate of ~ 2.1 large-scale rearrangements per million years after the divergence from a common cetartiodactyl ancestor ~ 60 Myr ago². This rate compares to ~ 1.9 rearrangements per million years within the primate lineage (Supplementary Table 11). A total of 20 and 18 cetartiodactyl EBRs (shared by pigs and cattle) were detected using the pig and human genomes as a reference, respectively.

Pig-specific EBRs were enriched for LTR endogenous retrovirus 1 (LTR-ERV1) transposons and satellite repeats (Supplementary Table 12), indicating that these two families of repetitive sequences have contributed to chromosomal evolution in the pig lineage. Different families of transposable elements seem to have been active in the cetartiodactyl ancestor. The cetartiodactyl EBRs are enriched for LINE1 elements and tRNA^{Glu}-derived SINEs. tRNA^{Glu}-derived SINEs, previously found over-represented in cetartiodactyl EBRs defined in the bovine genome¹⁰, originated in the common ancestor of cetartiodactyls²¹. Our observation that these elements are also enriched in porcine EBRs strongly supports the hypothesis that active transposable elements promote lineage-specific genomic rearrangements.

A stringent set of porcine to human one-to-one orthologues using the MetaCore database revealed that porcine EBRs and adjacent intervals are enriched for genes involved in sensory perception of taste ($P < 8.9 \times 10^{-6}$; FDR < 0.05), indicating that taste phenotypes may have been affected by events associated with genomic rearrangements. Pigs have a limited ability to taste NaCl²². *SCNN1B*, a gene encoding a sodium channel involved in the perception of salty tastes, is located in a porcine-specific EBR. Another gene, *ITPR3*, encoding a receptor for inositol triphosphate and a calcium channel involved in the perception of umami and sweet tastes, has been affected by the insertion of several porcine-specific SINE mobile elements into its 3' untranslated region (3' UTR), consistent with our observation of a higher density of transposable elements in EBRs. In addition to 8 bitter taste receptor genes annotated by Ensembl and which were used in the gene enrichment analysis, we identified 9 intact genes, to give a total number of 17 TAS2R receptors in the pig (Supplementary Table 13). This compares to 18 intact bitter taste receptors in cattle, 19 in horse, 15 in dog and 25 in humans^{23,24}. Of the 14 bitter taste receptor genes that were mapped to a specific pig chromosome (SSC), 10 were found near 2 EBRs on SSC5 and SSC18 (Supplementary Tables 13 and 15). We also found that at least four taste receptors (*TAS1R2*, *TAS2R1*, *TAS2R40* and *TAS2R39*) have been under relaxed selection (Supplementary Information). Pigs are not sensitive to bitter tastes and tolerate higher concentrations of bitter compounds than humans^{22,25}. Thus, pigs can eat food that is unpalatable to humans. A review of the porcine taste transduction network (Supplementary Fig. 17) revealed additional genes affected by rearrangements that affect apical and taste

receptor cell' processes. Together with the observed over-representation of genes related to 'adrenergic receptor activity' and 'angiotensin and other binding' categories in the pig EBRs (Supplementary Fig. 18), our data indicate that chromosomal rearrangements significantly contributed to adaptation in the suid lineage.

Population divergence and domestication

Divergence between Asian and European wild boar

We investigated the evolution within *Sus scrofa* in Eurasia by sequencing ten individual unrelated wild boars from different geographical areas. In total, 17,210,760 single nucleotide polymorphisms (SNPs) were identified among these ten wild boars. The number of SNPs segregating in the four Asian wild boars (11,472,192) was much higher than that observed in the six European wild boars (6,407,224) with only 2,212,288 shared SNPs. This higher nucleotide diversity was visible in the distribution of heterozygous sites of the Asian compared to the European wild boar genomes (Fig. 2). Phylogenomic analyses of complete genome sequences from these wild boars and six domestic pigs revealed distinct Asian and European lineages (Supplementary Fig. 23) that split during the mid-Pleistocene 1.6–0.8 Myr ago (Calabrian stage, Frantz, L. A. F. *et al.*, manuscript submitted). Colder climates during the Calabrian glacial intervals probably triggered isolation of populations across Eurasia. Admixture analyses (Supplementary Information) within Eurasian *Sus scrofa* disclosed gene flow between the northern Chinese and European populations consistent with pig migration across Eurasia, between Europe and northern China, throughout the Pleistocene. Our demographic analysis on the whole-genome sequences of European and Asian wild boars (Fig. 3) revealed an increase in the European population after pigs arrived from China. During the Last Glacial Maximum (LGM; ~20,000 years ago)²⁶, however, Asian and European populations both suffered population bottlenecks. The drop in population size was more pronounced in Europe than Asia (Fig. 3), suggesting a greater impact of the LGM in northern European regions and probably resulting in the observed lower genetic diversity in modern European wild boar.

The deep phylogenetic split between European and Asian wild boars is further supported by our observation of 1,272,737 fixed differences between the six European and four Asian wild boars, 1,706 of which are non-synonymous mutations in 1,191 different genes. Genes involved in sensory perception, immunity and host defence were among the most rapidly evolving genes (Supplementary Table 28), further strengthening the conclusions from our analysis of immunity-related pig proteins. This conclusion is further supported by our observation that these genes are also enriched in porcine segmental duplications (Supplementary Information).

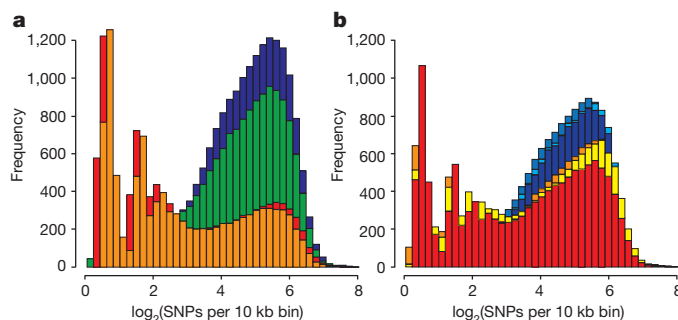


Figure 2 | Distribution of heterozygosity for individual pig genomes. Shown is the distribution of the heterozygosity as the \log_2 (SNPs) per 10k bin. **a**, Wild *Sus scrofa*: blue, south China; green, north China; orange, Italian; red, Dutch. **b**, Breeds: blue, Chinese breeds (Jiangquhai, Meishan, Xiang); red-yellow, European breeds (Hampshire, large white, landrace). Note that the Hampshire breed is a North American breed of European origin.

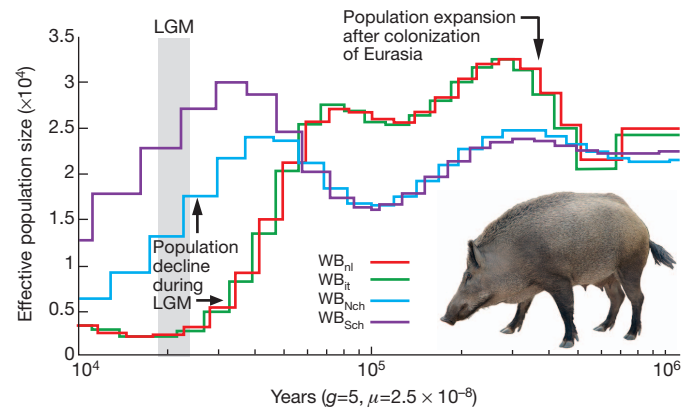


Figure 3 | Demographic history of wild boars. Demographic history was inferred using a hidden Markov model (HMM) approach as implemented in pairwise sequentially Markovian coalescence (PSMC)⁴⁵. In the absence of known mutation rates for pig, we used the default mutation rate for human (μ) of 2.5×10^{-8} . For the generation time (g) we used an estimate of 5 years. The Last Glacial Maximum (LGM) is highlighted in grey. WB_{Nl}, wild boar Netherlands; WB_{It}, wild boar Italy; WB_{Nch}, wild boar north China; WB_{Sch}, wild boar south China.

To investigate further whether specific regions in the genome of European and Asian wild boar have been under positive selection, a selective sweep analysis was performed on the ten wild boar genome sequences using an approach similar to that recently described in the comparison of Neanderthal and *Homo sapiens* genomes²⁷. Regions in the genome under strong positive selection after the divergence of these two populations are expected to share fewer derived alleles. Using stringent criteria (Supplementary Information), we identified a total of 251 putative selective sweep regions, with an average size of 111,269 base pairs (bp), together comprising around 1% of the genome and harbouring 365 annotated protein-coding genes (Supplementary Table 26). Many of these regions (112) do not contain any currently annotated protein-coding genes. In contrast, the 10 putative selective sweep regions located between positions 39–43 Mb on SSC3 together harbour 93 genes. This SSC3 region (Supplementary Fig. 25) is directly adjacent to the centromere and exhibits low recombination rates²⁸. Low recombining regions have been shown to be more prone to selective sweeps and meiotic drive^{29,30}. Although similar large putative selective sweep regions close to the centromere were only observed on SSC6 between positions 56.2–57.5 Mb, on most chromosomes selective sweep regions tended to cluster in the central part of chromosomes, thus exhibiting a clear correlation with regions of low recombination (Supplementary Fig. 27). As expected, regions with the highest nucleotide differentiation between European and Asian wild boars were observed in high recombination regions towards the end of the chromosomes on both metacentric and acrocentric chromosomes²⁸.

The putative selective sweep regions displayed significant over-representation of genes involved in RNA splicing and RNA processing, indicating possible changes in the regulation of genes at the level of RNA processing (Supplementary Table 27). Several of these genes (*CELF1*, *CELF6*, *WDR83*, *RBM39*, *RBM6*, *HNRNPA1*, *HNRNPM*) are involved in alternative splicing, and, small differences in expression might affect alternatively spliced transcripts of specific genes. Evolution of regulatory splicing factors such as heterogeneous ribonucleoprotein particle (hnRNP) proteins has been proposed as an evolutionary model for alternative splicing³¹, and genetic variation in such factors can affect alternative splicing and result in different phenotypes or disease³². Our observation that specific genes involved in splicing show accelerated evolution in the pig lineage (Fig. 1) supports this hypothesis. Of particular interest is the selective sweep region observed at position 26 Mb on SSC3 around the *ERI2* gene (Fig. 4), which encodes ERI1 exoribonuclease family member 2. Different gene variants have

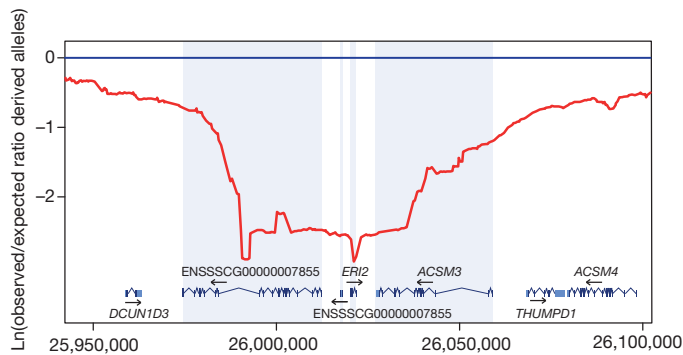


Figure 4 | Putative selective sweep region around the *ERI2* gene on SSC3. The y axis shows the log-transformed value of the ratio for the observed/expected derived allele frequency using a sliding window at a bin size of 50,000 bp. The x axis shows the position on SSC3 in base pairs.

been fixed in European and Asian wild boar coding for proteins that differ at two amino acid positions: Cys52Arg and His358Leu encoded by exons 3 and 9 of the *ERI2* gene, respectively. The precise function of *ERI2* is unknown but the *ERI1* exoribonuclease family members have been shown to be involved in mRNA decay³³ and in *Caenorhabditis elegans* *ERI-1* has been shown to be involved in the degradation of microRNAs (miRNAs)³⁴.

Independent domestication and admixture events in domestic breeds

A phylogenetic tree constructed using four European wild boar and domestic pigs and six East Asian wild boar and domestic pigs revealed a clear distinction between European and Asian breeds, thus substantiating the hypothesis that pigs were independently domesticated in western Eurasia and East Asia³. An admixture analysis revealed European influence in Asian breeds, and a ~35% Asian fraction in European breeds (Supplementary Table 24). These results are consistent with the known exchange of genetic material between European and Asian pig breeds³⁵. We also observed that European breeds form a paraphyletic clade, which cannot be solely explained by varying degrees of Asian admixture (Supplementary Information). Within each continent, our analysis revealed different degrees of relatedness between breeds and their respective wild relatives (Supplementary Table 20).

During domestication, pigs were often allowed to roam in a semi-managed state and recurrent admixture between wild and domesticated individuals was not uncommon, especially in Europe³⁵. Thus, the most likely explanation for the paraphyletic pattern seen in domestic individuals is a long history of genetic exchange between wild and domestic pigs.

The pig as a biomedical model

The pig is an important biomedical model and the ability to generate transgenics and knockouts in combination with somatic nuclear cloning procedures has resulted in a number of models for specific human diseases³⁶. Naturally occurring mutations also offer opportunities to use pigs as biomedical models^{37,38}. To explore the potential for natural models further, predicted porcine protein sequences were compared with their human orthologues. We observed 112 positions where the porcine protein has the same amino acid that is implicated in a human disease (Supplementary Table 29). Most of these changes in humans have been shown to increase risk in multifactorial traits such as obesity (*ADRB3*, *SDC3*) and diabetes (*PPPIRA*, *SLC30A8*, *ZNF615*) or shown to result in relatively mild phenotypes (for example, dyslexia: *KIAA0319*) or late-onset diseases such as Parkinson's disease (*LRRK2*, *SNCA*) and Alzheimer's disease (*TUBD1*, *BLMH*, *CEP192*, *PLAU*). These porcine variants are of interest, as they will allow detailed

characterization in an experimental model organism whose physiology is very similar to that of human.

Among 32,548 non-synonymous mutations identified by sequencing 48 individual pigs, representing 8 different European and Asian breeds and wild boars³⁹, 6 protein variants implicated in human disease were identified (Supplementary Table 30). In addition, another 157 nonsense mutations in 142 genes were identified, 11 of which have also been implicated in human disease (Supplementary Table 31). Most of these 11 variants were only observed in a heterozygous state and those for which homozygous individuals were observed probably result in either a mild phenotype (*ASS1*, mild form of citrullinaemia in humans) or in phenotypes unlikely to affect the fitness of wild boars (*RBBP8*, pancreatic carcinomas). Our estimate for the average number of nonsense mutations per individual (~30) is smaller than that observed in humans⁴⁰ despite the observed threefold higher nucleotide diversity in pigs³⁹. This is in agreement with the higher effective population size in the pig compared to that for the human population, which exhibited a strong bottleneck followed by an exponential increase in size during recent history⁴¹.

When considering pig-to-human xenotransplantation, porcine endogenous retroviruses (PERVs) pose a risk of zoonotic infection. The pig genome contains fewer endogenous retroviruses than many vertebrates, including humans and mice, and most PERVs were characterized as defective. However, the potential risk posed by reactivation of rare replication-competent PERVs and defective PERVs by recombination remains, as shown for murine ERVs (XMRV)⁴². Most PERVs consist of γ and γ -like groups (68%), with β -retroviral ERVs comprising the second most abundant group (Supplementary Fig. 15). Our phylogenetic study shows a particularly close relationship between the most intact $\gamma 1$ group of PERVs ($\gamma 1$) and murine γ -ERVs, suggesting a potential recent instance of murine-to-porcine transmission of $\gamma 1$ ERVs (Supplementary Fig. 15). We identified 20 almost intact PERV $\gamma 1$ loci (Supplementary Table 10), none of which contained a complete set of *gag*, *pol* or *env* open reading frames, indicating that these proviruses are not replicable. We also identified four β -retroviral PERVs, each containing defects, primarily in *env*. These were distantly related to intracisternal type A particle (IAP) proviruses of mice and the mouse mammary tumour virus (MMTV)-like (HML) proviruses of humans. None of the above loci was shared in more than 120 pigs tested, indicating considerable PERV polymorphisms.

Conclusion

The draft pig genome sequence reported here has illuminated the evolution of *Sus scrofa* and confirmed its speciation in South East Asia and subsequent domestication at multiple regions across Eurasia. The high-quality annotated reference genome sequence has already proven to be a critical framework for comparing individual genomes^{39,43,44} and its value is further illustrated in associated papers published elsewhere (<http://www.biomedcentral.com/series/swine>). The genome sequence also provides a valuable resource enabling effective uses of pigs both in agricultural production and in biomedical research.

METHODS SUMMARY

Assembly. We constructed a hybrid *de novo* assembly based primarily on sequences from BAC clones sequenced clone-by-clone and supplemented with Illumina whole-genome shotgun (WGS) reads. BAC clones were selected from the high-resolution physical (BAC contig) map⁶ with CHORI-242 library clones prepared from DNA from a single Duroc sow (Duroc 2-14) chosen preferentially. The WGS sequence data were generated using DNA isolated from the same animal. The BAC-derived sequence data were assembled into sequence contigs using Phrap on a clone-by-clone basis and subsequently independently assembled WGS contigs (Supplementary Information) were used to extend BAC clone-derived sequence contigs and to close gaps between clone-derived contigs. Further details and other methods are described in Supplementary Information.

Received 16 June; accepted 27 September 2012.

1. Kumar, S. & Hedges, S. B. A molecular timescale for vertebrate evolution. *Nature* **392**, 917–920 (1998).
2. Meredith, R. W. *et al.* Impacts of the Cretaceous Terrestrial Revolution and KPg extinction on mammal diversification. *Science* **334**, 521–524 (2011).
3. Larson, G. *et al.* Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science* **307**, 1618–1621 (2005).
4. Schook, L. B. *et al.* Swine Genome Sequencing Consortium (SGSC): a strategic roadmap for sequencing the pig genome. *Comp. Funct. Genomics* **6**, 251–255 (2005).
5. Archibald, A. L. *et al.* Pig genome sequence – analysis and publication strategy. *BMC Genomics* **11**, 438 (2010).
6. Humphray, S. J. *et al.* A high utility integrated map of the pig genome. *Genome Biol.* **8**, R139 (2007).
7. Flicke, P. *et al.* Ensembl 2012. *Nucleic Acids Res.* **40**, D84–D90 (2012).
8. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4**, 44–57 (2009).
9. Barreiro, L. B. & Quintana-Murci, L. From evolutionary genetics to human immunology: how selection shapes host defense genes. *Nature Rev. Genet.* **11**, 17–30 (2010).
10. Bovine Genome Sequencing and Analysis Consortium. The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science* **324**, 522–528 (2009).
11. Vandenbroeck, K. *et al.* Gene sequence, cDNA construction, expression in *Escherichia coli* and genetically approached purification of porcine interleukin-1 beta. *Eur. J. Biochem.* **217**, 45–52 (1993).
12. Renard, C. *et al.* The genomic sequence and analysis of the swine major histocompatibility complex. *Genomics* **88**, 96–110 (2006).
13. Tanaka-Matsuda, M., Ando, A., Rogel-Gaillard, C., Chardon, P. & Uenishi, H. Difference in number of loci of swine leukocyte antigen classical class I genes among haplotypes. *Genomics* **93**, 261–273 (2009).
14. Schwartz, J. C., Lefranc, M. P. & Murtaugh, M. P. Evolution of the porcine (*Sus scrofa domestica*) immunoglobulin κ locus through germline gene conversion. *Immunogenetics* **64**, 303–311 (2012).
15. Schwartz, J. C., Lefranc, M. P. & Murtaugh, M. P. Organization, complexity and allelic diversity of the porcine (*Sus scrofa domestica*) immunoglobulin lambda locus. *Immunogenetics* **64**, 399–407 (2012).
16. Uenishi, H. *et al.* Genomic structure around joining segments and constant regions of swine T-cell receptor α/δ (TRA/TRD) locus. *Immunology* **109**, 515–526 (2003).
17. Uenishi, H. *et al.* Genomic sequence encoding diversity segments of the pig TCR δ chain gene demonstrates productivity of highly diversified repertoire. *Mol. Immunol.* **46**, 1212–1221 (2009).
18. Eguchi-Ogawa, T., Toki, D. & Uenishi, H. Genomic structure of the whole D-J-C clusters and the upstream region coding V segments of the TRB locus in pig. *Dev. Comp. Immunol.* **33**, 1111–1119 (2009).
19. Sambrook, J. G. *et al.* Identification of a single killer immunoglobulin-like receptor (KIR) gene in the porcine leukocyte receptor complex on chromosome 6q. *Immunogenetics* **58**, 481–486 (2006).
20. Nguyen, D. T. *et al.* The complete swine olfactory subgenome: expansion of olfactory receptor gene repertoire in the pig genome. *BMC Genomics* (in the press).
21. Shimamura, M., Abe, H., Nikaido, M., Ohshima, K. & Okada, N. Genealogy of families of SINEs in cetaceans and artiodactyls: the presence of a huge superfamily of tRNA(Glu)-derived families of SINEs. *Mol. Biol. Evol.* **16**, 1046–1060 (1999).
22. Hellekant, G. & Danilova, V. Taste in domestic pig, *Sus scrofa*. *J. Anim. Physiol. Anim. Nutr. (Berl.)* **82**, 8–24 (1999).
23. Fischer, A., Gilad, Y., Man, O. & Pääbo, S. Evolution of bitter taste receptors in humans and apes. *Mol. Biol. Evol.* **22**, 432–436 (2005).
24. Dong, D., Jones, G. & Zhang, S. Dynamic evolution of bitter taste receptor genes in vertebrates. *BMC Evol. Biol.* **9**, 12 (2009).
25. Nelson, S. L. & Sanregret, J. D. Response of pigs to bitter-tasting compounds. *Chem. Senses* **22**, 129–132 (1997).
26. Yokoyama, Y., Lambek, K., De Deckker, P., Johnston, P. & Fifield, L. K. Timing of the Last Glacial Maximum. *Nature* **406**, 713–716 (2000).
27. Green, R. E. *et al.* A draft sequence of the Neandertal genome. *Science* **328**, 710–722 (2010).
28. Tortereau, F. *et al.* Sex specific differences in recombination rate in the pig are correlated with GC content. *BMC Genomics* (in the press).
29. Barton, N. H. Genetic hitchhiking. *Phil. Trans. R. Soc. Lond. B* **355**, 1553–1562 (2000).
30. Lytle, T. W. Cheaters sometimes prosper: distortion of mendelian segregation by meiotic drive. *Trends Genet.* **9**, 205–210 (1993).
31. Ast, G. How did alternative splicing evolve? *Nature Rev. Genet.* **5**, 773–782 (2004).
32. Garcia-Blanco, M. A., Baraniak, A. P. & Lasda, E. L. Alternative splicing in disease and therapy. *Nature Biotechnol.* **22**, 535–546 (2004).
33. Kupsc, J. M., Wu, M.-J., Marzluff, W. F., Thapar, R. & Duronio, R. J. Genetic and biochemical characterization of *Drosophila* Snipper: A promiscuous member of the metazoan 3' hExo/ERI-1 family of 3' to 5' exonucleases. *RNA* **12**, 2103–2117 (2006).
34. Kennedy, S., Wang, D. & Ruvkun, G. A conserved siRNA-degrading RNase negatively regulates RNA interference in *C. elegans*. *Nature* **427**, 645–649 (2004).
35. White, S. From globalized pig breeds to capitalist pigs: a study in animals cultures and evolutionary history. *Environ. Hist.* **16**, 94–120 (2011).
36. Walters, E. M. *et al.* Completion of the swine genome will simplify the production of swine as a large animal biomedical model. *BMC Med. Genomics* (in the press).
37. Gillard, E. F. *et al.* A substitution of cysteine for arginine 614 in the ryanodine receptor is potentially causative of human malignant hyperthermia. *Genomics* **11**, 751–755 (1991).
38. Murgiano, L., Tammen, I., Harlizius, B. & Drögemüller, C. A *de novo* germline mutation in MYH7 causes a progressive dominant myopathy in pigs. *BMC Genomics* (in the press).
39. Bosse, M. *et al.* Regions of homozygosity in the porcine genome: Consequence of demography and the recombination landscape. *PLoS Genet.* **8**, e1003100 (2012).
40. MacArthur, D. G. *et al.* A systematic survey of loss-of-function variants in human protein-coding genes. *Science* **335**, 823–828 (2012).
41. Keinan, A. & Clark, A. G. Recent explosive human population growth has resulted in an excess of rare variants. *Science* **336**, 740–743 (2012).
42. Paprotka, T. *et al.* Recombinant origin of the retrovirus XMRV. *Science* **333**, 97–101 (2011).
43. Fang, X. *et al.* The sequence and analysis of an inbred pig genome. *GigaScience* (in the press).
44. Rubin, C. J. *et al.* Strong signatures of selection in the domestic pig genome. *Proc. Natl. Acad. Sci. USA* (in the press).
45. Li, H. & Durbin, R. Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493–496 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements The authors recognize the contributions of the following individuals towards the establishment of the Swine Genome Sequencing Consortium and their leadership in realizing this effort: J. Jen, P. J. Burfening, D. Hamernik, R. A. Easter, N. Merchen, R. D. Green, J. Cassady, B. Harlizius, M. Boggess and M. Stratton. Also the authors acknowledge A. Hernandez, C. Wright at the University of Illinois Keck Center for Comparative and Functional Genomics; N. Bruneau and Prof. Ning Li for their contribution to PERV studies; D. Goodband and D. Berman for their efforts in genome annotation; D. Grafham of the Wellcome Trust Sanger Institute for his efforts in the genome assembly and J. Hedegaard, M. Nielsen and R. O. Nielsen for their contribution on the miRNA analysis. We also recognize contributions from the National Institute of Agrobiological Sciences and the Institute of Japan Association for Techno-innovation in Agriculture, Forestry and Fisheries, Tsukuba, Japan, H. Shinkai, T. Eguchi-Ogawa, K. Suzuki, D. Toki, T. Matsumoto, N. Fujishima-Kanaya, A. Mikawa, N. Okumura, M. Tanaka-Matsuda, K. Kurita, H. Sasaki, K. Kamiya, T. Bito and N. Fujitsuka. We acknowledge support from the USDA CSREES/NIFA Swine Genome Coordination Program, College of Agricultural, Consumer and Environmental Sciences, University of Illinois; College of Agriculture and Life Sciences, Iowa State University; North Carolina Agricultural Research Service; USA National Pork Board; Iowa Pork Producers Association; North Carolina Pork Council; Danish government; TOPIGS Research Center IPG The Netherlands; INRA Genoscope, France; Wellcome Trust Sanger Institute and BGI. We are grateful to the genome team at NCBI for their assistance in checking the Sscrofa10.2 assembly and for their independent annotation of the sequence. This project was also partially supported by grants: BBSRC grants (Ensembl): BB/E010520/1, BB/E010520/2, BB/E025328/1; EC FP6 'Cutting edge genomics for sustainable animal breeding (SABRE)'; EC FP7 'Quantomics'; C. J. Martin Overseas Based Biomedical Fellowship from the Australian NHMRC (575585); BBSRC (BB/H005935/1); Next-Generation BioGreen 21 Program (PJ009019, PJ0081162012), RDA, Republic of Korea; Consolider programme from Ministry of Research (Spain); NIH R13 RR020283A; NIH R13 RR032267A; ILLU 535-314; ILLU 538-379; ILLU-538-312; ILLU-538-34; CSREES, NIFA for funding genome coordination activities; NIH grant 5 P41LM006252; MAFF grants (RPPIAUGT-AG 1101/1201); USDA-NRI-2009-35205-05192; USDA-NRSP8 Bioinformatics Coordination and Pig Genome Coordination funds; US-UK Fulbright Commission; Next-Generation BioGreen 21 (no. PJ0080892011) Program, RDA, Republic of Korea; USDA-ARS Project Plan 1235-51000-055-00D; USDA-ARS Project Plan 1265-32000-098-00D; USDA-NRI-2006-35204-17337 USDA AFRI NIFA/DHS 2010-39559-21860; NIH P20-RR017686; USDA ARS; USDA-NRSP8 Bioinformatics; USDA ARS Beltsville Area Summer Undergraduate Fellowships; BBSRC grant BB/G004013/1; NSFC Outstanding Youth grant (31025026); The Swedish Research Council FORMAS; The Swedish Wenner-Gren Foundation; European Commission FP6 funded project LSHB-CT-2006-037377; BioGreen21, RDA grant PJ00622901; BioGreen21, RDA grant PJ00622902; BioGreen21, RDA grant PJ00622903; BioGreen21, RDA grant PJ00622903. The research leading to these results has received funding from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement no. 249894 (SelSweep); NIH P20-RR017686; NIH NIDA P30 DA018310; NIH NIDA R21 DA027548; NIAS, RDA grant PJ001758; BioGreen21, RDA grant PJ006229; NIAS, RDA grant 20040301034467; ANR grant ANR07-GANI-001 Delisus; Danish funding agencies: FTP/DFP (09-066598); DSF/Strategic Growth Technologies (09-067036); the Lundbeck foundation (374/06); DCSC (Scientific Computing); The Funds for International Cooperation from the Ministry of Science and Technology of China 2002AA229061; PL-Grid project: POIG.02.03.00-00-007/08-00 'Genome Assembly'; USDA-NIFA-CREES AG 2006-35216-16668; AG 2002-34480-11828; AG 2003-34480-13172; AG 2004-34480-14417; AG 2005-34480-15939; AG 2006-34480-17150; AG 2008-34480-19328; AG 2009-34480-19875; AG 2002-35205-12712; somatic cell genomics: Integrating QTL Discovery and Validation; AG 2008-35205-18769; AG 2009-65205-05642; AG 2004-3881-02193; AG 2011-67015-30229; AG 58-5438-2-313; AG 58-5438-7-317; and AG 58-0208-7-149; NIH grant 5 P41 LM006252.

Author Contributions Manuscript main text: A.L.A., M.A.M.G., L.B.S., H.U., C.K.T., Y.T., M.F.R., C.P., S.L., D.M., H.-J.M., D.M.L., H.K.I., L.A.F.F., G.L.M.C.; project coordination: A.L.A., M.A.M.G., L.B.S., M.F.R., D.M., J.R., C.Chu., H.U., M.C., K.E.; project initiation: A.L.A.,

M.A.M.G., L.B.S., M.F.R., D.M., M.F., C.W.B., P.C., G.A.R., M.Y., J.R., L.B.; library preparation and sequencing: S.J.H., C.S., C.C.I., S.M., L.M., M.J., Y.Lu, X.X., P.N., Jia.Z., G.Z., A.L.A., R.C.C., T.M., H.Ka., K.-T.L., T.-H.K., H.-S.P., E.-W.P., J.-H.K., S.-H.C., S.-J.O., Ji.W., Ju.W., J.-T.J.; genome assembly: A.L.A., M.C., S.L., C.S., P.D., H.-J.M., H.U., D.M., B.S., T.F., Y.Li, N.D., R.R.-G., R.L., K.H., W.C.; repetitive DNA analysis: G.J.F. (leader), J.J., F.DeS., H.-J.M.; gene content and genome evolution: S.F., B.L.A., S.W., S.S.; conservation of synten and evolutionary breakpoints: D.M.L. (leader), J.N., L.A., B.C., H.A.L., J.M., J.K., D.K.G., K.E.F.; speciation: L.A.F.F., M.A.M.G., O.M., H.-J.M., J.G.S.; divergence of Asian and European wild boar: H.-J.M., M.Bo., M.A.M.G., L.A.F.F.; annotation: S.S., B.L.A., T.M., C.K.T., Y.S., M.By., R.C., J.R., E.F., Z.-L.H., W.L., M.P.-E.; RNA analysis: O.M., R.P.M.A.C., H.U., C.A., H.T., B.T., P.S., M.F., J.G., C.B., F.P., H.H., Z.B., J.F.; neuropeptides: J.V.S., B.R.S., S.R.-Z.; pig domestication: L.A.F.F., R.P.M.A.C., H.-J.M., M.Bo., S.O., G.L., L.R., J.G.S.; population admixture: L.A.F.F., J.G.S.; biomedical models: B.D., L.R., K.S., M.A.M.G.; immune response: C.K.T., (co-leader) C.R.-G. (co-leader), H.D.D., J.E.L., A.A., B.B., J.S., D.B., F.B., M.By., S.B., C.Che., D.C.-S., R.C., E.F., E.G., J.G.R.G., J.L.H., T.H., Z.-L.H., R.K., J.K.L., K.M., M.P.M., T.M., G.P., J.M.R., J.S., H.U., Jie Z., S.Z.; olfactory and taste receptor analysis: C.P. (leader), D.T.N., K.L.; dN/dS analysis: H.Ki. (leader), H.A., K.-W.K.; PERV and retroviral insertions: C.R.-G., A.H., P.J., J.B., G.S., L.S., R.W., Y.T. (leader); segmental duplications: O.M., Y.P., Z.-Q.D., M.F.R.

Author Information The final assembly (Sscrofa10.2) has been deposited in the public sequence databases (GenBank/EMBL/DBJ) under accession number AEMK01000000. The primary source of the Sscrofa10.2 assembly is the NCBI ftp site (http://ftp.ncbi.nih.gov/genbank/genomes/Eukaryotes/vertebrates_mammals/Sus_scrofa/Sscrofa10.2/). The chromosomes are CM000812–CM00830 and CM001155. They are built from 5,343 placed scaffolds, with GenBank accession numbers GL878569–GL882503 and JH114391–JH118402. The 4,562 unplaced scaffolds of Sscrofa10.2 have accessions in the ranges GL892100–GL896682 and JH118403–JH118999. Illumina sequences for the sequenced wild boars and individuals of the other breeds, aligned against build10.2, have been deposited in the European Nucleotide Archive (ENA) under project number ERP001813. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.A.M.G. (martien.groenen@wur.nl) or A.L.A. (alan.archibald@roslin.ed.ac.uk). This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported licence. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-sa/3.0/>

Martien A. M. Groenen^{1*}, Alan L. Archibald^{2*}, Hirohide Uenishi³, Christopher K. Tuggle⁴, Yasuhiro Takeuchi⁵, Max F. Rothschild⁴, Claire Rogel-Gaillard⁶, Chankyu Park⁷, Denis Milan⁸, Hendrik-Jan Megens¹, Shengting Li^{9,10}, Denis M. Larkin¹¹, Heebal Kim¹², Laurent A. F. Frantz¹, Mario Caccamo¹³, Hyeonju Ahn¹², Bronwen L. Aken¹⁴, Anna Anselmo¹⁵, Christian Anthon¹⁶, Loretta Auvil¹⁷, Bouabid Badaoui¹⁵, Craig W. Beattie¹⁸, Christian Bendixen¹⁹, Daniel Berman²⁰, Frank Blecha²¹, Jonas Blomberg²², Lars Bolund^{9,10}, Mirte Bosse¹, Sara Botti¹⁵, Zhan Buje¹⁹, Megan Bystrom⁴, Boris Capitanu¹⁷, Denise Carvalho-Silva²³, Patrick Chardon⁶, Celine Chen²⁴, Ryan Cheng⁴, Sang-Haeng Choi²⁵, William Chow¹⁴, Richard C. Clark¹⁴, Christopher Clee¹⁴, Richard P. M. A. Crooijmans¹, Harry D. Dawson²⁴, Patrice Dehaes⁸, Fioravante De Sapio²⁶, Bert Dibbitts¹, Nizar Drou¹³, Zhi-Qiang Du⁴, Kellye Eversole²⁶, João Fadista^{19†}, Susan Fairley¹⁴, Thomas Faraut³, Geoffrey J. Faulkner[†], Katie E. Fowler²⁷, Merete Fredholm¹⁶, Eric Fritz⁴, James G. R. Gilbert¹⁴, Elisabetta Giuffra^{5,15}, Jan Gorodkin¹⁶, Darren K. Griffin²⁷, Jennifer L. Harrow¹⁴, Alexander Hayward²⁸, Kerstin Howe¹⁴, Zhi-Liang Hu⁴, Sean J. Humphray^{14†}, Toby Hunt¹⁴, Henrik Hornshøj¹⁹, Jin-Tae Jeon²⁹, Patric Jern²⁸, Matthew Jones¹⁴, Jerzy Jurka³⁰, Hiroyuki Kanamori^{3,31}, Ronan Kapetanovic², Jaebum Kim^{7,32}, Jae-Hwan Kim³³, Kyu-Won Kim³⁴, Tae-Hun Kim³⁵, Greger Larson³⁶, Kyooyeol Lee³, Kyung-Tai Lee³⁵, Richard Leggett¹³, Harris A. Lewin³⁷, Yingrui Li⁹, Wansheng Liu³⁸, Jane E. Loveland¹⁴, Yao Lu⁹, Joan K. Lunney²⁰, Jian Ma³⁹, Ole Madsen¹, Katherine Mann^{20†}, Lucy Matthews¹⁴, Stuart McLaren¹⁴, Takeya Morozumi³¹, Michael P. Murtaugh⁴⁰, Jitendra Narayan¹¹, Dinh Truong Nguyen⁷, Peixiang Ni⁹, Song-Jung Oh⁴¹, Suneel Onteru⁴, Frank Panitz¹⁹, Eung-Woo Park³⁵, Hong-Seog Park²⁵, Geraldine Pascal⁴², Yogesh Paudel¹, Miguel Perez-Enciso⁴³, Ricardo Ramirez-Gonzalez¹³, James M. Reecy⁴, Sandra Rodriguez-Zas⁴⁴, Gary A. Rohrer⁴⁵, Lauretta Rund⁴⁴, Yongming Sang²¹, Kyle Schachtschneider⁴⁴, Joshua G. Schraiber⁴⁶, John Schwartz⁴⁰, Linda Scobie⁴⁷, Carol Scott¹⁴, Stephen Searle¹⁴, Bertrand Servin⁸, Bruce R. Southey⁴⁴, Goran Sperber⁴⁸, Peter Stadler⁴⁹, Jonathan V. Sweedler⁵⁰, Hakim Tafer⁴⁹, Bo Thomsen¹⁹, Rashmi Wali⁴⁷, Jian Wang⁹, Jun Wang^{9,51}, Simon White¹⁴, Xun Xu⁹, Martine Yerle⁸, Guojie Zhang^{9,52}, Jianguo Zhang⁹, Jie Zhang⁵³, Shuhong Zhao⁵³, Jane Rogers¹³, Carol Churcher¹⁴ & Lawrence B. Schook⁵⁴

¹Animal Breeding and Genomics Centre, Wageningen University, De Elst 1, 6708 WD, Wageningen, The Netherlands. ²The Roslin Institute and R(D)SVS, University of Edinburgh, Easter Bush, Midlothian EH25 9RG, UK. ³National Institute of Agrobiological Sciences, 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan. ⁴Department of Animal Science and Center for Integrated Animal Genomics, Iowa State University, 2255 Kildee Hall, Ames 50011, USA. ⁵MRC/UCL Centre for Medical Molecular Virology and Wohl Virion Centre, Division of Infection & Immunity, University College London, Crutwick

Building, Gower Street, London WC1E 6BT, UK. ⁶INRA, Laboratory of Animal Genetics and Integrative Biology/AgroParisTech, Laboratory of Animal Genetics and Integrative Biology/CEA, DSV, IRCM, Laboratoire de Radiobiologie et Etude du Génome, Domaine de Vilvert, F-78350 Jouy-en-Josas, France. ⁷Department of Animal Biotechnology, Konkuk University, 1 Hwayang-dong, Kwangjin-gu, Seoul 143-701, South Korea. ⁸INRA, Laboratoire de Génétique Cellulaire, Chemin de Borde-Rouge, Auzeville, 31320 Castanet Tolosan, France. ⁹BGI-Shenzhen, Shenzhen 518083, China. ¹⁰Department of Biomedicine, Aarhus University, DK-8000 Aarhus C, Denmark. ¹¹Institute of Biological, Environmental and Rural Sciences, Aberystwyth University, Penglais Campus, Aberystwyth, Ceredigion SY23 3DA, UK. ¹²Department of Agricultural Biotechnology and C&K Genomics, Seoul National University, Gwanakgu, Seoul 151-742, South Korea. ¹³The Genome Analysis Centre, Norwich Research Park, Norwich NR4 7UH, UK. ¹⁴Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SA, UK. ¹⁵Parco Tecnologico Padano, Via Einstein, Loc. C. Codazza, 26900 Lodi, Italy. ¹⁶Center for non-coding RNA in Technology and Health, IBHV University of Copenhagen, Frederiksberg, Denmark. ¹⁷Illinois Informatics Institute, University of Illinois, Urbana, Illinois 61801, USA. ¹⁸Department of Surgery, University of Illinois, Chicago, Illinois 60612, USA. ¹⁹Department of Molecular Biology and Genetics, Aarhus University, DK-8830 Tjele, Denmark. ²⁰USDA ARS BARC Animal Parasitic Diseases Laboratory, Beltsville, Maryland 20705, USA. ²¹Department of Anatomy and Physiology, College of Veterinary Medicine, Kansas State University, Manhattan, Kansas 66506, USA. ²²Clinical Virology, Department of Medical Sciences, Uppsala University, Building D1, Academic Hospital, 751 85 Uppsala, Sweden. ²³European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SD, UK. ²⁴Diet, Genomics, Immunology Laboratory, Beltsville Human Nutrition Research Center, United States Department of Agriculture, BARC-East 10300 Baltimore Ave Beltsville, Maryland 20705, USA. ²⁵Korean Research Institute of Bioscience and Biotechnology, 125 Gwahak ro, Yuseong gu, Daejeon 305-806, South Korea. ²⁶Eversole Associates and the Alliance for Animal Genome Research, 5207 Wyoming Road, Bethesda, Maryland 20816, USA. ²⁷School of Biosciences, The University of Kent, Giles Lane, Canterbury, Kent CT2 7NJ, UK. ²⁸Science for Life Laboratory, Department of Medical Biochemistry and Microbiology, Uppsala University, BMC, Box 582, SE75123 Uppsala, Sweden. ²⁹Department of Animal Sciences, College of Agriculture and Life Sciences, Gyeongsang National University, Jinju 660-701, South Korea. ³⁰Genetic Information Research Institute, 1925 Landings Drive, Mountain View, California 94043, USA. ³¹Institute of Japan Association for Techno-innovation in Agriculture, Forestry and Fisheries, 446-1 Ippaizuka, Kamiyokoba, Tsukuba, Ibaraki 305-0854, Japan. ³²Institute for Genomic Biology, University of Illinois, Urbana, Illinois 61801, USA. ³³Animal Genetic Resources Station, National Institute of Animal Science, RDA, San 4, Yongsanri, Ulsan, Namwon 590-832, South Korea. ³⁴C&K Genomics, Gwanakgu, Seoul 151-742, South Korea. ³⁵Animal Genomics and Bioinformatics Division, National Institute of Animal Science, RDA, 77 Chuksan gil, Wonsun gu, Suwon 441-706, South Korea. ³⁶Durham Evolution and Ancient DNA, Department of Archaeology, Durham University, Durham DH1 3LE, UK. ³⁷Department of Evolution and Ecology, The UC Davis Genome Center, University of California, Davis, California 95618, USA. ³⁸Department of Dairy and Animal Sciences, Center for Reproductive Biology and Health (CRBH), College of Agricultural Sciences, The Pennsylvania State University, 305 Henning Building, University Park, Pennsylvania 16802, USA. ³⁹Department of Bioengineering and Institute for Genomic Biology, University of Illinois, Urbana, Illinois 61801, USA. ⁴⁰Department of Veterinary and Biomedical Sciences, University of Minnesota, 1971 Commonwealth Avenue, St Paul, Minnesota 55108, USA. ⁴¹Jeju National University, 102 Jejudaehakno, Jeju 690-756, South Korea. ⁴²INRA UMR85/CNRS UMR7247 Physiologie de la Reproduction et des Comportements/IFCE, F-37380 Nouzilly, France and Université François Rabelais de Tours, F-37041 Tours, France. ⁴³CREA, Centre for Research in Agricultural Genomics (CRAG) and Facultat de Veterinaria UAB, Campus Universitat Autònoma Barcelona, Bellaterra E-08193, Spain. ⁴⁴Department of Animal Sciences, University of Illinois, Urbana, Illinois 61801, USA. ⁴⁵USDA, ARS, US Meat Animal Research Center, Clay Center, Nebraska 68933, USA. ⁴⁶Department of Integrative Biology, University of California, Berkeley, California 94720-3140, USA. ⁴⁷Department of Life Sciences, Glasgow Caledonian University, Glasgow G4 0BA, UK. ⁴⁸Department of Neuroscience, Biomedical Center, Uppsala University, PO Box 593, 751 24 Uppsala, Sweden. ⁴⁹Bioinformatics Group, Department of Computer Science, Interdisciplinary Center for Bioinformatics, Universität Leipzig, Leipzig, Germany. ⁵⁰Department of Chemistry, University of Illinois, Urbana, Illinois 61801, USA. ⁵¹Novo Nordisk Foundation Center for Basic Metabolic Research and Department of Biology, University of Copenhagen, DK-2200 Copenhagen, Denmark. ⁵²BGI-Europe, DK-2200 Copenhagen N, Denmark. ⁵³Key Lab of Animal Genetics, Breeding, and Reproduction of Ministry Education, Huazhong Agricultural University, Wuhan 430070 PR China, Huazhong Agricultural University, Wuhan 430070, China. ⁵⁴Department of Animal Sciences and Institute for Genomic Biology, University of Illinois, Urbana, Illinois 61801, USA. [†]Present addresses: Lund University Diabetes Centre, CRC, Malmö University Hospital, SE-205 02 Malmö, Sweden (J.F.); Mater Medical Research Institute, and School of Biomedical Sciences, University of Queensland, Brisbane, 4072 Queensland, Australia (G.J.F.); Illumina Inc. Chesterford Research Park, Little Chesterford, Nr Saffron Walden, Essex CB10 1XL, UK (S.J.H.); Department of Molecular Microbiology, Washington University School of Medicine, Saint Louis, Missouri 63110, USA (K.M.).

*These authors contributed equally to this work.

‡Deceased.

Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes

A list of authors and their affiliations appears at the end of the paper

Pancreatic cancer is a highly lethal malignancy with few effective therapies. We performed exome sequencing and copy number analysis to define genomic aberrations in a prospectively accrued clinical cohort ($n = 142$) of early (stage I and II) sporadic pancreatic ductal adenocarcinoma. Detailed analysis of 99 informative tumours identified substantial heterogeneity with 2,016 non-silent mutations and 1,628 copy-number variations. We define 16 significantly mutated genes, reaffirming known mutations (*KRAS*, *TP53*, *CDKN2A*, *SMAD4*, *MLL3*, *TGFBR2*, *ARID1A* and *SF3B1*), and uncover novel mutated genes including additional genes involved in chromatin modification (*EPC1* and *ARID2*), DNA damage repair (*ATM*) and other mechanisms (*ZIM2*, *MAP2K4*, *NALCN*, *SLC16A4* and *MAGEA6*). Integrative analysis with *in vitro* functional data and animal models provided supportive evidence for potential roles for these genetic aberrations in carcinogenesis. Pathway-based analysis of recurrently mutated genes recapitulated clustering in core signalling pathways in pancreatic ductal adenocarcinoma, and identified new mutated genes in each pathway. We also identified frequent and diverse somatic aberrations in genes described traditionally as embryonic regulators of axon guidance, particularly *SLIT/ROBO* signalling, which was also evident in murine Sleeping Beauty transposon-mediated somatic mutagenesis models of pancreatic cancer, providing further supportive evidence for the potential involvement of axon guidance genes in pancreatic carcinogenesis.

Pancreatic cancer is the fourth leading cause of cancer death, with an overall 5-year survival rate of $<5\%$, statistics that have not changed in almost 50 years¹. Advances in neoadjuvant and adjuvant chemotherapeutic regimens have resulted in some improvement in outcome, but pancreatectomy remains the single most effective treatment modality for pancreatic cancer, and offers the only potential for cure. Only 20% of patients present with localized, non-metastatic disease which is suitable for resection². Those who undergo resection and receive adjuvant therapy have a median survival of 12–22 months and a 5-year survival of 20–25%³. Existing systemic therapies are only modestly effective and the median survival for patients with metastatic disease remains 6 months. Genomic characterization of pancreatic ductal adenocarcinoma (PDAC), which accounts for over 90% of pancreatic cancer, has so far focused on targeted polymerase chain reaction (PCR)-based exome sequencing of primary and metastatic lesions propagated as xenografts or cell lines⁴. A deeper understanding of the underlying molecular pathophysiology of the clinical disease is needed to advance the development of effective therapeutic and early detection strategies.

Clinical cohort

A cohort of 142 consecutive patients with primary operable, untreated PDAC who underwent pancreatectomy with curative intent (pre-operative clinical stages I and II) were recruited, and consent was obtained for genomic sequencing through the Australian Pancreatic Cancer Genome Initiative (APGI), the Baylor College of Medicine Pancreatic Cancer Genome Project and the Ontario Institute for Cancer Research Pancreatic Cancer Genome Study (ABO collaboration) between June 2005 and June 2011 as part of the International Cancer Genome Consortium (ICGC)⁵. Detailed clinico-pathological characteristics of the cohort demonstrated features typical of resected PDAC with regard to tumour size, grade, lymph node metastasis and survival when compared to multiple retrospectively acquired cohorts^{6–8}, defining the accrued population as representative of the clinical disease in the community (Supplementary Table 1 and Supplementary Fig. 1).

Cellularity and mutation detection

A major challenge in genomic sequencing is the low malignant epithelial cell content of many cancers, which can adversely impact on the sensitivity of mutation detection. Most sequencing studies so far have used samples with $>70\%$ tumour cellularity, or cell lines/xenografts^{4,9}. To implement genomic sequencing approaches in clinical practice, it is imperative to efficiently and accurately detect actionable mutations in diagnostic clinical samples. We devised methodologies to overcome the challenges associated with extensive desmoplastic stroma that is characteristic of the majority of PDAC, and these strategies facilitated the discovery of novel molecular mechanisms in the pathophysiology of this disease. The cellularity of each primary sample was estimated through pathological review, deep amplicon-based sequencing of exons 2 and 3 of *KRAS* (average depth of $1,000\times$), and single nucleotide polymorphism (SNP) array-based cellularity estimates using a novel algorithm (qpure)¹⁰. *KRAS* mutations were identified in 93% of 142 cases and tumour cellularity ranged from 5% to 85% with a mean of 38% (Supplementary Table 2, Supplementary Figs 2 and 3, and Supplementary Methods).

To inform cellularity thresholds for subsequent analyses, we defined the impact of stromal DNA content on mutation detection by exome capturing and sequencing different mixtures of cancer cell line and matched germline DNA (100%, 80%, 60%, 40%, 20% and 10% cell line DNA) when sequenced to a depth of $70\times$ coverage. Using these data as a standard, the median sensitivity to detect true positives across all samples in the cohort with greater than 20% epithelial cellularity was estimated at 45% (Supplementary Table 3). An informative cohort of 99 patients who had greater than 20% cellularity and/or ≥ 10 validated somatic mutations was taken forward for further analysis.

Mutation detection and CNV analysis

We performed hybrid-selection-based capture and sequencing of the entire exomes of tumour and matched normal DNA derived from all 142 patients using a combination of capture systems and next-generation sequencing platforms (see Supplementary Methods). The

Table 1 | Mutations in pancreatic ductal adenocarcinoma (*n* = 99)

Mutation class	Total
Missense	1,684
Nonsense	99
Splice site	89
Insertion/deletion	144
Non-silent	2,016
Silent	611

sequence depths at each site (APGI 65×, BCM 104× and OICR 205×) were adopted to ensure suitable sensitivity across their respective cohorts (Supplementary Table 3). In the informative 99 samples, we detected 2,627 high-confidence mutations, 2,016 of which were non-silent (Table 1). A total of 1,502 of these events (1,350 non-silent) were independently validated via an orthogonal sequencing method (see Supplementary Methods). The average number of mutations detected per patient was 26 (range 1–116), consistent with the expected sensitivity based on cellularity estimates and previous studies^{4,11} (Supplementary Table 2). We confirmed the high prevalence of genetic aberrations known to be important in PDAC and observed mutations in 38 of the 79 genes (48% overlap) that occurred more than once previously reported by ref. 4, and 186 of all 998 mutated genes (19% overlap) in that study. We also defined a large number of novel mutations (1,456 genes), most of which occurred at low frequency (see Supplementary Tables 4–6 and Supplementary Fig. 4 for detailed comparisons). The observed transversion/transition rates in the cohort correlated closely with those previously reported in PDAC cell lines and xenografts (Supplementary Table 7).

Significant mutated gene analysis¹² of genes with non-silent mutations that occurred in 2 or more individual cancers identified 16 genes in the top 20 mutated genes in 2 of 3 stringent analytical approaches (Table 2, Supplementary Table 8 and Supplementary Methods) and reaffirmed the importance of mutations known to occur in PDAC: *KRAS*, *TP53*, *CDKN2A*, *SMAD4*, *MLL3*, *TGFBR2*, *ARID1A* and *SF3B1*. Novel significantly mutated genes included additional genes involved in chromatin modification (*EPC1* and *ARID2*) and *ATM*, recently implicated as a PDAC susceptibility gene through bi-allelic inactivation in a case of familial PDAC (germline mutation and loss of heterozygosity (LOH) in the tumour)¹³. Aberrations of *ATM* occurred in 8% of our cohort (mutated in 5%, LOH or loss in 5%, with two patients exhibiting both mutation and LOH or loss) and mutations detected in other genes not previously reported: *ZIM2*, *MAP2K4*, *NALCN*, *SLC16A4* and *MAGEA6* (Table 2). GISTIC2.0¹⁴ identified 30 genes affected by copy-number alterations (Q value <0.0001) and included losses of *CDKN2A* and *SMAD4* (Supplementary Table 4).

Table 2 | Significantly mutated genes in pancreatic ductal adenocarcinoma

Gene symbol	Gene name and protein function	SB mutagenesis*	shRNA†
<i>KRAS</i>	Oncogene; GTPase; activation of MAPK activity	Yes	Yes
<i>TP53</i>	Tumour suppressor p53; DNA damage response	–	Yes
<i>CDKN2A</i>	Cyclin-dependent kinase inhibitor 2A; G1/S transition of mitotic cell cycle; tumour suppressor	Yes	–
<i>SMAD4</i>	Mothers against decapentaplegic homologue 4; BMP signalling pathway	Yes	Yes
<i>MLL3</i>	Myeloid/lymphoid or mixed-lineage leukaemia protein 3; DNA binding; regulation of transcription	Yes	Yes
<i>TGFBR2</i>	Transforming growth factor-β receptor type II; regulation of growth	Yes	–
<i>ARID1A</i>	AT-rich interactive domain-containing protein 1A; SWI/SNF complex; chromatin modification	Yes	Yes
<i>ARID2</i>	AT-rich interactive domain-containing protein 2; chromatin modification	Yes	–
<i>EPC1</i>	Enhancer of polycomb homologue 1; histone acetylation	Yes	–
<i>ATM</i>	Ataxia telangiectasia mutated; DNA damage response	–	Yes
<i>SF3B1</i>	Splicing factor 3B subunit 1; nuclear mRNA splicing	–	Yes
<i>ZIM2</i>	Zinc finger imprinted 2; regulation of transcription	–	Yes
<i>MAP2K4</i>	Dual specificity mitogen-activated protein kinase kinase 4; Toll-like receptor signalling pathway	Yes	Yes
<i>NALCN</i>	Sodium leak channel non-selective protein; sodium channel activity	–	Yes
<i>SLC16A4</i>	Solute carrier family 16 member 4; monocarboxylate transporter	–	Yes
<i>MAGEA6</i>	Melanoma-associated antigen 6; protein binding	–	ND

ND, not determined.

* Significant insertion sites in two independent Sleeping Beauty mutagenesis screens^{17,18}.

† *In vitro* shRNA screens in 102 cancer cell lines with effect on cell survival¹⁹.

Pathways in pancreatic cancer

To better understand potential underlying mechanisms of importance in PDAC, we performed a series of pathway analyses using genes that were recurrently mutated in two or more individuals using GeneGO¹⁵, and identified mechanisms known to be important in cancer: G1/S checkpoint machinery ($P = 1.49 \times 10^{-3}$), apoptosis ($P = 1.32 \times 10^{-4}$), regulation of angiogenesis ($P = 7.72 \times 10^{-4}$) and TGF-β signalling ($P = 9.50 \times 10^{-4}$). Interestingly, novel gene signatures were enriched in our cohort, including axon guidance ($P = 5.30 \times 10^{-5}$) (Supplementary Table 9). The inclusion of mutation data for 24 cases from ref. 4 strengthened the association of axon guidance ($P = 3.3 \times 10^{-7}$), and was more evident still when all mutated genes in our data set were used as input ($P = 4.67 \times 10^{-8}$).

Functional relevance of genomic events

Differentiating somatic driving events of carcinogenesis from passenger mutations is a major challenge in cancer genomics¹⁶. Despite significant advances in computational algorithms, experimental evidence of functional relevance is paramount. We used data from three published experimental biological screens to infer functional consequences for the individual genomic events and the pathways we identified. These included data from two independent Sleeping Beauty transposon (SB) mutagenesis screens in *Kras* transgenic mouse models of PDAC^{17,18} and an *in vitro* short hairpin RNA (shRNA) screen which examined the consequences of downregulating 11,194 putative cancer genes on survival in a panel of 102 cell lines (13 pancreatic)¹⁹ (Supplementary Methods and Supplementary Figs 5 and 6). Data from these screens confirmed the functional importance of *KRAS*, *TP53*, *CDKN2A* and *SMAD4* mutations and attributed potential functional relevance to most significantly mutated genes—*MLL3*, *TGFBR2*, *SF3B1*, *EPC1*, *ARID1A*, *ARID2*, *MAP2K4*, *ATM*, *NALCN*, *ZIM2*, *SLC16A4* (Table 2)—and many genes mutated at low frequency (Supplementary Table 4).

Pathway analysis of high confidence insertions in SB transposon mutagenesis screens demonstrated enrichment for axon guidance genes ($P = 1.6 \times 10^{-3}$), providing independent supportive evidence for a potential role in the pathogenesis of PDAC. In these screens, 14 genes involved in axon guidance pathways were detected (5 genes common to both). In addition, a further 32 genes were mutated in at least one SB pancreatic tumour (out of 21) but did not meet the significance threshold with the stringent analyses that were applied¹⁷ (Supplementary Tables 10 and 11).

Axon guidance pathway genes

The class of genes traditionally described for their roles in axon guidance (semaphorins, slits, netrins and ephrins) are important regulators of

normal neuronal migration and positioning during embryonic development. More recently, they have been implicated in cancer cell growth, survival, invasion and angiogenesis²⁰; however, the incidence of aberrations in these genes in cancer is largely unknown. We identified recurrent mutations and copy-number variations (CNVs) of axon guidance pathway genes in this cohort (Fig. 1 and Supplementary Table 4): *SLIT2* and *ROBO2* mutations were present in 5% of patients, with focal copy-number losses of *ROBO1*, and *SLIT2* detected by GISTIC2.0 analysis and confirmed by manual review, potentially having an impact on a further 15% of the cohort, suggesting that aberrant SLIT/ROBO signalling is potentially a common feature of PDAC (Figs 1 and 2). In addition, we used targeted PCR-based sequencing of an additional 30 cases of PDAC for axon guidance genes and identified mutations in *ROBO1* in two patients and additional mutations in *SLIT2* and *ROBO2* (one patient each). Low mRNA expression of the *ROBO2* receptor was associated with poor patient survival ($P = 0.04$). Furthermore, high mRNA expression of *ROBO3*, a known inhibitor of *ROBO2* signalling²¹, demonstrated an appropriate reciprocal inverse association with poor survival ($P < 0.006$) (Fig. 2).

Class 3 semaphorins (*SEMA3A* and *SEMA3E*) exhibited significant amplification in 18% of patients and an additional 3% harboured mutations (Fig. 1). Semaphorins signal through neuropilin and plexin receptors to elicit their effects²². *SEMA3A* amplification correlated with high mRNA expression on microarray ($P = 0.03$), and high mRNA expression of *SEMA3A* and *PLXNA1*, another molecule central to semaphorin signalling, were both associated with poor patient survival on univariate analysis (Fig. 3a), and were independently prognostic on multivariate analyses with clinico-pathological variables (Supplementary Table 12).

To elucidate further the significance of the observed CNV events, we reviewed copy number, CNV segment size and changes in heterozygosity of axon guidance genes in a recent independent CNV analysis of 39 fine-needle aspiration biopsies²³ and the 16 PDAC cell lines in the

CONAN database (<http://www.sanger.ac.uk/cosmic>)²⁴. Overall, the predominant changes recapitulated our studies, showing frequent focal losses within genes involved in SLIT/ROBO signalling, and gains in genes involved in canonical semaphorin signalling (Supplementary Tables 4, 13 and 14).

To assess whether dysregulation of axon guidance genes is associated with early neoplastic transformation, as are many developmental signalling pathways, we examined mRNA expression in murine models of early pancreatic carcinogenesis (*in vitro* acinar-to-ductal metaplasia and *in vivo* pancreatic injury). Expression levels of components of SLIT/ROBO and semaphorin signalling changed progressively from normal pancreas, through acinar-to-ductal metaplasia and pancreatic injury to genetically engineered murine PDAC, indicating a role for the dysregulation of these axon guidance genes in tumour initiation and progression (Fig. 3b and Supplementary Table 15).

Discussion

We devised methodologies to optimize mutation detection for clinical samples in a large cohort of patients and reaffirm known mutations in PDAC, better define their prevalence in a large cohort of early PDAC, and identify potential novel drivers in this disease. Somatic mutations in *ATM* were identified in a significant proportion of patients (8%), highlighting the importance of BRCA-mediated DNA damage repair mechanisms in sporadic PDAC as well as familial disease¹³. Previously, mutations in individual genes involved in chromatin remodelling such as *ARID1A*²⁵ have been described and additional genes identified here (*EPC1* and *ARID2*) infer that chromatin remodelling may have an important role in PDAC, along with other cancer types²⁶.

Novel mutations in genes traditionally described for their roles in axon guidance were also observed by a combination of genomic data and supportive experimental evidence from independent murine SB mutagenesis screens. Axon guidance is integral to organogenesis,

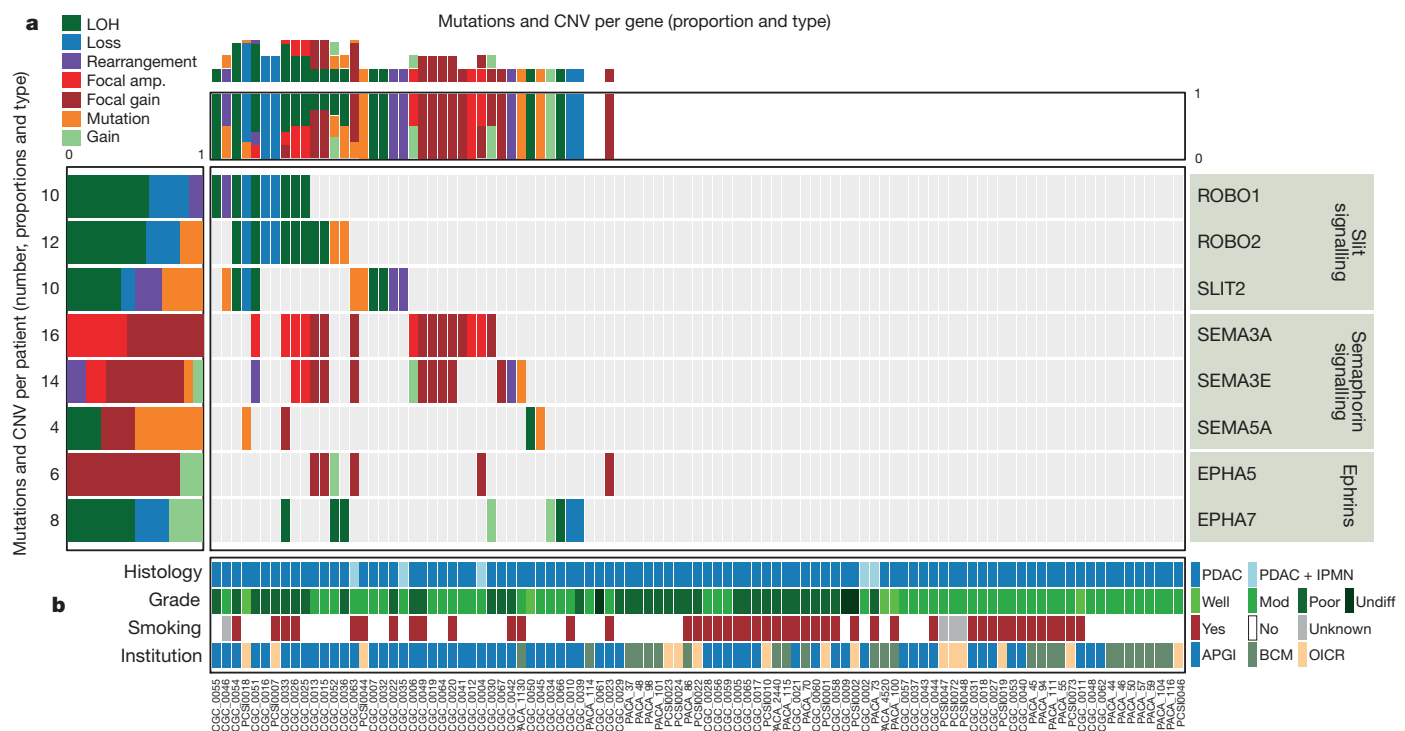


Figure 1 | Mutations and copy number variation in axon guidance genes. Axon guidance pathway genes with recurrent mutations and/or copy-number changes defined by GISTIC2.0 analysis ($Q < 0.2$), and manually reviewed for focal alterations. **a**, SNV and CNV frequency per patient with gene-centric summary (left) and patient-centric summary (top); numbers of patients with mutations and proportion of each event are presented. Please see

Supplementary Table 4 for further details. **b**, Clinico-pathological variables for individual patients. APGI, Australian Pancreatic Cancer Genome Initiative; BCM, Baylor College of Medicine; IPMN, intraductal papillary mucinous neoplasm; Mod, moderately differentiated; OICR, Ontario Institute for Cancer Research; PDAC, pancreatic ductal adenocarcinoma; Undiff, undifferentiated.

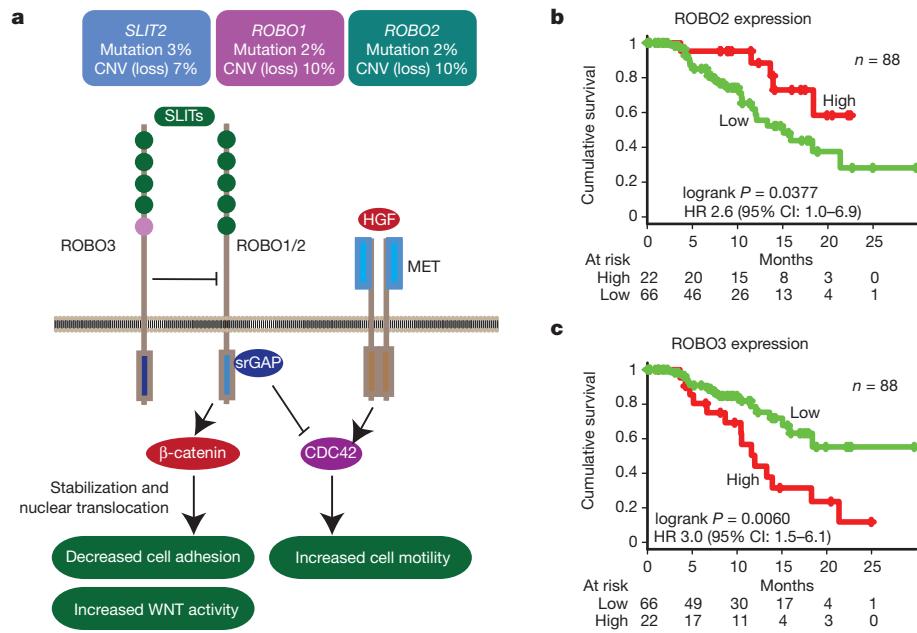


Figure 2 | SLIT/ROBO signalling in pancreatic ductal adenocarcinoma. **a**, SLIT/ROBO signalling normally enhances β -catenin complex formation with E-cadherin and suppresses WNT signalling activity. Loss of ROBO1/2 signalling promotes stabilization of β -catenin, which decreases E-cadherin complex formation and cell adhesion and augments WNT signalling activity through increased nuclear translocation of β -catenin. In addition, SLIT/ROBO signalling can downregulate MET signalling activity; loss of ROBO signalling activity promotes MET signalling downstream and may have an impact on

therapeutic strategies aimed at inhibiting MET activity at the receptor level. (Adapted from ref. 20.) Aberrations in *SLIT2* and/or *ROBO1/2* affected 23% of patients (6% mutated with 1 patient showing mutations in both *SLIT2* and *ROBO2*), with 18% demonstrating CNV corresponding to loss of the gene. **b**, **c**, High expression of SLIT receptor ROBO2 was associated with a better prognosis (**b**), and high expression of ROBO3, an inhibitor of ROBO2, showed an inverse relationship, with high levels associated with poor survival (**c**). HR, hazard ratio.

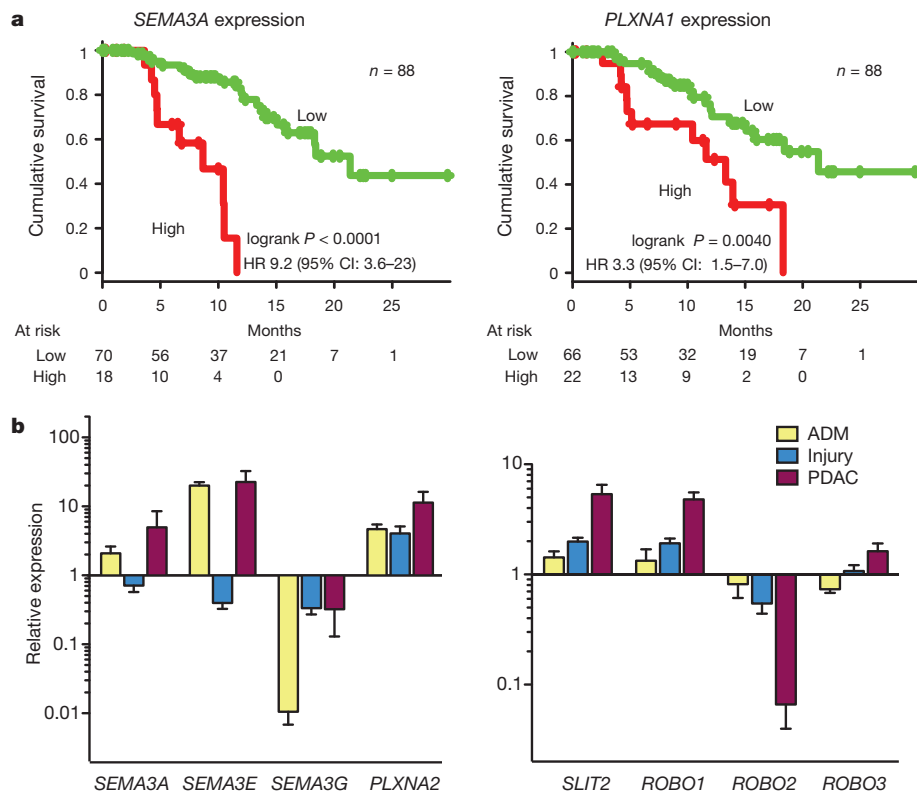


Figure 3 | Axon guidance genes in human and murine pancreatic ductal adenocarcinoma. **a**, Kaplan–Meier survival curves showing co-segregation of aberrant expression of components of semaphorin signalling with outcome. Amplification at *SEMA3A* and *PLXNA1* loci was associated with high mRNA expression and both are independent poor prognostic factors. **b**, Quantitative RT–PCR for components of semaphorin and SLIT/ROBO signalling in murine

models of early (acinar-to-ductal metaplasia (ADM) and pancreatic injury) and established PDAC in genetically engineered mice with a *Pdx1*-promoter-driven activating mutation of *Kras* and mutant *Tp53* allele (*Pdx1-Cre*; *LSL-Kras*^{G12D}; *LSL-Trp53*^{R172H}). Error bars represent standard error of the mean (see Supplementary Table 15 for details).

regeneration, wound healing and other basic cellular processes^{22,27}. The widespread genomic aberrations observed here in axon guidance genes suggests that they may have a role in PDAC, joining mounting evidence in other cancers^{20,28}, including a recent report demonstrating *ROBO2* mutations in liver-fluke-associated cholangiocarcinoma²⁹. In addition, evidence from cancers of the lung, breast, kidney and cervix implicate aberrant SLIT/ROBO signalling in carcinogenesis²⁰; *Robo1* knockout mice develop bronchial hyperplasia and focal dysplasia, and inactivation of *Slit2* and *Slit3* leads to the development of hyperplastic disorganized lesions in the breast²⁰. Upregulation of MET and WNT signalling have important roles in PDAC, and recent data indicate that SLIT/ROBO signalling modulates MET and WNT signalling activity through CDC42 and β -catenin, respectively²⁰. Loss of SLIT/ROBO signalling can potentially be an alternative mechanism for deregulating these pathways downstream of their receptors, and in addition could influence the activity of inhibitors that target these upstream components, for example, MET inhibitors (Fig. 2).

Class 3 semaphorins are the only secreted semaphorins in vertebrates. They regulate cell growth, invasiveness and angiogenesis, and are highly expressed in metastatic cells in many cancer types^{30,31}. Although aberrant semaphorin signalling in cancer seems to be organ specific³², our finding that high expression of *SEMA3A* and its receptor *PLXNA1* co-segregates with poor patient survival is supported by a previous study that reported this association and also demonstrated promotion of invasiveness of PDAC cell lines by *SEMA3A*³¹. Therapeutics targeting molecules involved in axon guidance have been developed as potential strategies to facilitate neuronal regeneration after injury³³, but are yet to be assessed for their role in cancer treatment.

As illustrated here, global genomic analysis of large, well-annotated and clinically homogeneous cohorts of patients can identify mechanisms that are common among genomically diverse cancers, and will be pivotal in the development of novel therapeutic strategies that are guided by the determination of the molecular phenotype of individual patients³⁴. Future work will be required to determine which key components, when damaged, drive the disease, and these mechanisms will need to be assessed in molecularly well-characterized preclinical models³⁵. The potential therapeutic strategies identified will then require testing in appropriate clinical trials that are specifically designed to target subsets of patients stratified according to well-defined molecular markers^{36,37}.

METHODS SUMMARY

Sample acquisition and processing. Samples used were prospectively acquired and restricted to primary operable, non-pretreated pancreatic ductal adenocarcinoma. Representative sections were reviewed independently by at least one additional pathologist with specific expertise in pancreatic diseases. Samples either had full face frozen sectioning performed in optimal cutting temperature (OCT) medium, or the ends excised and processed in formalin to verify the presence of carcinoma in the sample to be sequenced and to estimate the percentage of malignant epithelial nuclei in the sample relative to stromal nuclei. Macrodissection was performed if required to excise areas that did not contain malignant epithelium. **Sequencing.** Cellularity of each tumour sample was estimated with pathology review, deep sequencing of *KRAS* and a method developed using genome-wide SNP array data (qpure¹⁰). Exon capture was performed using the SureSelect II or Nimblegen capture methods and paired-end sequenced on the SOLiD (v4) or GAI/HiSeq platforms. Somatic mutations were called and then verified on the Ion Torrent Personal Genome Machine (Life Technologies Corporation) and 454 (Hoffman-La Roche Limited).

Analysis. Significantly mutated genes were identified using the Genome MuSiC package¹². DNA copy number analyses were performed using the Illumina HumanOmni1 Quad genotyping arrays and GenoCN software. Recurrent and significant copy number changes were identified using GISTIC2.0¹⁴. Functional enrichment of gene categories was assessed using the Metacore package (Thomson-Reuters Corporation) and the MSigDB v3.0 database³⁸. All sample information and data for mutation, copy number and expression analyses were submitted to the ICGC DCC at <http://dcc.icgc.org/>. A complete description of the materials and methods including approvals for human research and animal experimentation is provided in Supplementary Information.

Received 9 January; accepted 4 September 2012.

Published online 24 October; corrected online 14 November 2012 (see full-text HTML version for details).

- Jemal, A., Siegel, R., Xu, J. & Ward, E. Cancer statistics, 2010. *CA Cancer J. Clin.* **60**, 277–300 (2010).
- Butturini, G. *et al.* Influence of resection margins and treatment on survival in patients with pancreatic cancer: meta-analysis of randomized controlled trials. *Arch. Surg.* **143**, 75–83 (2008).
- Neoptolemos, J. P. *et al.* Adjuvant chemotherapy with fluorouracil plus folinic acid vs gemcitabine following pancreatic cancer resection: a randomized controlled trial. *J. Am. Med. Assoc.* **304**, 1073–1081 (2010).
- Jones, S. *et al.* Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* **321**, 1801–1806 (2008).
- International Cancer Genome Consortium. International network of cancer genome projects. *Nature* **464**, 993–998 (2010).
- Biankin, A. V. *et al.* Expression of S100A2 calcium-binding protein predicts response to pancreatectomy for pancreatic cancer. *Gastroenterology* **137**, 558–568 (2009).
- Chang, D. K. *et al.* Margin clearance and outcome in resected pancreatic cancer. *J. Clin. Oncol.* **27**, 2855–2862 (2009).
- Jamieson, N. B. *et al.* A prospective comparison of the prognostic value of tumor- and patient-related factors in patients undergoing potentially curative surgery for pancreatic ductal adenocarcinoma. *Ann. Surg. Oncol.* **18**, 2318–2328 (2011).
- Wang, L. *et al.* Whole-exome sequencing of human pancreatic cancers and characterization of genomic instability caused by MLH1 haploinsufficiency and complete deficiency. *Genome Res.* **22**, 208–219 (2012).
- Song, S. *et al.* qpure: A tool to estimate tumor cellularity from genome-wide single-nucleotide polymorphism profiles. *PLoS ONE* **7**, e45835 (2012).
- Samuel, N. & Hudson, T. J. The molecular and cellular heterogeneity of pancreatic ductal adenocarcinoma. *Nature Rev. Gastroenterol. Hepatol.* **9**, 77–87 (2012).
- Dees, N. D. *et al.* MuSiC: Identifying mutational significance in cancer genomes. *Genome Res.* **22**, 1589–1598 (2012).
- Roberts, N. J. *et al.* ATM mutations in patients with hereditary pancreatic cancer. *Cancer Discov.* **2**, 41–46 (2011).
- Mermel, C. H. *et al.* GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* **12**, R41 (2011).
- Sun, W. *et al.* Integrated study of copy number states and genotype calls using high-density SNP arrays. *Nucleic Acids Res.* **37**, 5365–5377 (2009).
- Campbell, P. J. *et al.* The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature* **467**, 1109–1113 (2010).
- Mann, K. M. *et al.* Sleeping Beauty mutagenesis reveals cooperating mutations and pathways in pancreatic adenocarcinoma. *Proc. Natl Acad. Sci. USA* **109**, 5934–5941 (2012).
- Pérez-Mancera, P. A. *et al.* The deubiquitinase USP9X suppresses pancreatic ductal adenocarcinoma. *Nature* **486**, 266–270 (2012).
- Cheung, H. W. *et al.* Systematic investigation of genetic vulnerabilities across cancer cell lines reveals lineage-specific dependencies in ovarian cancer. *Proc. Natl Acad. Sci. USA* **108**, 12372–12377 (2011).
- Mehlen, P., Delloye-Bourgeois, C. & Chédotal, A. Novel roles for Slits and netrins: axon guidance cues as anticancer targets? *Nature Rev. Cancer* **11**, 188–197 (2011).
- Sabatier, C. *et al.* The divergent Robo family protein rig-1/Robo3 is a negative regulator of slit responsiveness required for midline crossing by commissural axons. *Cell* **117**, 157–169 (2004).
- Trusolino, L. & Comoglio, P. M. Scatter-factor and semaphorin receptors: cell signalling for invasive growth. *Nature Rev. Cancer* **2**, 289–300 (2002).
- Birnbaum, D. J. *et al.* Genome profiling of pancreatic adenocarcinoma. *Genes Chromosomes. Cancer* **50**, 456–465 (2011).
- Bamford, S. *et al.* The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. *Br. J. Cancer* **91**, 355–358 (2004).
- Jones, S. *et al.* Somatic mutations in the chromatin remodeling gene *ARID1A* occur in several tumor types. *Hum. Mutat.* **33**, 100–103 (2012).
- Varela, I. *et al.* Exome sequencing identifies frequent mutation of the SWI/SNF complex gene *PBRM1* in renal carcinoma. *Nature* **469**, 539–542 (2011).
- Comoglio, P. M. & Trusolino, L. Invasive growth: from development to metastasis. *J. Clin. Invest.* **109**, 857–862 (2002).
- Chédotal, A., Kerjan, G. & Moreau-Fauvarque, C. The brain within the tumor: new roles for axon guidance molecules in cancers. *Cell Death Differ.* **12**, 1044–1056 (2005).
- Ong, C. K. *et al.* Exome sequencing of liver fluke-associated cholangiocarcinoma. *Nature Genet.* **44**, 690–693 (2012).
- Capparucci, L. & Tamagnone, L. Semaphorin signaling in cancer cells and in cells of the tumor microenvironment—two sides of a coin. *J. Cell Sci.* **122**, 1723–1736 (2009).
- Müller, M. W. *et al.* Association of axon guidance factor semaphorin 3A with poor outcome in pancreatic cancer. *Int. J. Cancer* **121**, 2421–2433 (2007).
- Ellis, L. M. The role of neuropilins in cancer. *Mol. Cancer Ther.* **5**, 1099–1107 (2006).
- Kikuchi, K. *et al.* *In vitro* and *in vivo* characterization of a novel semaphorin 3A inhibitor, SM-216289 or xanthohumol. *J. Biol. Chem.* **278**, 42985–42991 (2003).
- Cao, Y., DePinho, R. A., Ernst, M. & Voudsen, K. Cancer research: past, present and future. *Nature Rev. Cancer* **11**, 749–754 (2011).
- Pajic, M., Scarlett, C. J., Chang, D. K., Sutherland, R. L. & Biankin, A. V. Preclinical strategies to define predictive biomarkers for therapeutically relevant cancer subtypes. *Hum. Genet.* **130**, 93–101 (2011).
- Biankin, A. V. & Hudson, T. J. Somatic variation and cancer: therapies lost in the mix. *Hum. Genet.* **130**, 79–91 (2011).

37. Kris, M. G., Meropol, N. J. & Winer, E. P. (eds). *Accelerating Progress Against Cancer: ASCO's Blueprint for Transforming Clinical and Translational Cancer Research* (Am. Soc. Clin. Oncol., 2011).
38. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).

Supplementary Information is available in the online version of the paper.

Acknowledgements This paper is dedicated to Robert L. Sutherland who died on 10 October 2012 of pancreatic cancer. We would like to thank C. Axford, D. Gwynne, M.-A. Brancato, S. Rowe, M. Thomas, S. Simpson and G. Hammond for central coordination of the Australian Pancreatic Cancer Genome Initiative, data management and quality control; M. Martyn-Smith, L. Braatvedt, H. Tang, V. Papangelis and M. Beilin for biospecimen acquisition; and W. Waterson, J. Shepperd, E. Campbell and E. Glasov for their efforts at the Queensland Centre for Medical Genomics. We also thank M. B. Hodgins, M. Debeljak and D. Trusty for technical assistance at Johns Hopkins University, and J. Lau, M. Karaus, K. Rabe, L. Zhang and T. Smyrk at the Mayo Clinic. We acknowledge the following funding support: National Health and Medical Research Council of Australia (NHMRC; 631701, 535903, 427601, 535914); Australian Government: Department of Innovation, Industry, Science, Research and Tertiary Education (DIISRTE); Australian Cancer Research Foundation (ACRF); Queensland Government (NIRAP); University of Queensland; Cancer Council NSW (SRP06-01; ICGC09-01; SRP11-01); Cancer Institute NSW (06/ECF/1-24, 09/CFD/2-40, 07/CFD/1-03, 10/CRF/1-01, 08/RSA/1-15, 07/CFD/1-28, 10/CFD/2-26, 10/FRL/2-03, 06/RSA/1-05, 09/RIG/1-02, 10/TPG/1-04, 11/REG/1-10, 11/CFD/3-26); Garvan Institute of Medical Research; Avner Nahmani Pancreatic Cancer Research Foundation; R.T. Hall Trust; Petre Foundation; Jane Hemstitch in memory of Philip Hemstitch; Gastroenterological Society of Australia (GESA); American Association for Cancer Research (AACR) Landon Foundation – INNOVATOR Award; Royal Australasian College of Surgeons (RACS); Royal Australasian College of Physicians (RACP); Royal College of Pathologists of Australasia (RCPA); HGSC-BCM: NHGRI U54 HG003273; CPRIT grant RP101353-P7 (Tumor Banking for Genomic Research and Clinical Translation Site 1); The Ontario Institute for Cancer Research; The Ontario Ministry of Economic Development and Innovation; Canada Foundation for Innovation; Pancreatic Cancer Genetic Epidemiology Consortium, NIH grant R01 CA97075; The Agency for Science, Technology, and Research (Singapore); University of Verona and Italian Ministry of University (FIRB RBAP10AHJB), Rome, Italy; Cancer Research UK; Wellcome Trust; CPRIT (Cancer Prevention Research Institute of Texas); NIH P50CA062924 (SPORE) and P01CA134292 (PPG); The Sol Goldman Pancreatic Cancer Research Center; NCI grant P50 CA102701 (Mayo Clinic SPORE in Pancreatic Cancer) and NCI grant R01 CA97075 (Pancreatic Cancer Genetic Epidemiology Consortium); NIH SPORE grant 2P50CA101955 (UMN/UAB), and AIRC (Associazione Italiana Ricerca sul Cancro) 5xmille grant 12182, Italy.

Author Contributions The research network comprising the Australian Pancreatic Cancer Genome Initiative, the Baylor College of Medicine Cancer Genome Project and the Ontario Institute for Cancer Research Pancreatic Cancer Genome Study (ABO collaboration) contributed collectively to this study as part of the International Cancer Genome Consortium. Biospecimens were collected at affiliated hospitals and processed at each biospecimen core resource centre. Data generation and analyses were performed by the genome sequencing centres, cancer genome characterization centres and genome data analysis centres. Investigator contributions are as follows: S.M.G., A.V.B., J.V.P., R.L.S., R.A.G., D.A.W., M.-C.G., J.D.M., L.D.S. and T.J.H. (project leaders); A.V.B., S.M.G. and R.L.S. (writing team); A.L.J., J.V.P., P.J.W., J.L.F., C.L., M.A., O.H., J.G.R., D.T., C.X., S.W.G., F.N., S.S.o., G.K. and W.K. (bioinformatics/databases); D.K.M., I.H., S.L., C.N., S.M., A.Chr., T.Br., S.Wa., E.N., B.B.G., D.M.M., Y.Q.W., Y.H., L.R.L., H.D., R.E.D., R.S.M. and M.W. (sequencing); N.W., K.S.K., J.V.P., A.-M.P., K.N., N.C., M.G., P.J.W., M.J.C., M.P., J.W., N.K., F.Z., J.D., K.C., C.J.B., L.B.M., D.P., R.E.D., R.D.B., T.Be. and C.K.Y. (mutation, copy number and gene expression analysis); A.L.J., D.K.C., M.D.J., M.P., C.J.S., E.K.C., C.T., A.M.N., E.S.H., V.T.C., L.A.C., E.N., J.S.S., J.L.H., C.T., N.B. and M.Sc. (sample processing and quality control); A.J.G., J.G.K., R.H.H., C.A.I.-D., A.Cho., A.Mai., J.R.E., P.C. and A.S. (pathology assessment); J.W., M.J.C., M.P., C.K.Y. and mutation analysis team (network/pathway analysis and functional data integration); K.M.M., N.A.J., N.G.C., P.A.P.-M., D.J.A., D.A.L., L.F.A.W., A.G.R., D.A.T., R.J.D., I.R., A.V.P., E.A.M., R.L.S., R.H.H. and A.Maw. (functional screens); E.N., A.L.J., J.S.S., A.J.G., J.G.K., N.D.M., A.B., K.E., N.Q.N., N.Z., W.E.F., F.C.B., S.E.H., G.E.A., L.M., L.T., M.Sam., K.B., A.B., D.P., A.P., N.B., R.D.B., R.E.D., C.Y., S.Se., N.O., D.M., M.-S.T., P.A.S., G.M.P., S.G., L.D.S., C.A.I.-D., R.D.S., C.L.W., R.A.M., R.T.L., S.B., V.C., M.Sca., C.B., M.A.T., G.T., A.S. and J.R.E. (sample collection and clinical annotation); D.K.C., M.P., C.J.S., E.S.H., J.A.L., R.J.D., A.V.P. and I.R. (preclinical models).

Author Information BAM files and associated metadata in XML format have been uploaded to the European Genome-phenome Archive (EGA; <http://www.ebi.ac.uk/ega>) under accession numbers EGAS00001000154 and EGAS00001000343. Additional sequence data is located at dbGAP accession number phs000516.v1.p1. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.M.G. (s.grimmond@imb.uq.edu.au).

Andrew V. Biankin^{1,2,3}, Nicola Waddell⁴, Karin S. Kassahn⁴, Marie-Claude Gingras^{5,6}, Lakshmi B. Muthuswamy⁷, Amber L. Johns¹, David K. Miller⁴, Peter J. Wilson⁴, Ann-Marie Patch⁴, Jianmin Wu¹, David K. Chang^{1,2,3}, Mark J. Cowley¹, Brooke B. Gardiner⁴, Sarah Song⁴, Ivon Harliwong⁴, Senel Idrisoglu⁴, Craig Nourse⁴, Ehsan Nourbakhsh⁴, Suzanne Manning⁴, Shivangi Wani⁴, Milena Gongora⁴, Marina Pajic¹,

Christopher J. Scarlett^{1,8}, Anthony J. Gill^{1,9,10}, Andreia V. Pinho¹, Ilse Rooman¹, Matthew Anderson⁴, Oliver Holmes⁴, Conrad Leonard⁴, Darrin Taylor⁴, Scott Wood⁴, Qingying Xu⁴, Katia Nones⁴, J. Lynn Fink⁴, Angelika Christ⁴, Tim Bruxner⁴, Nicole Cloonan⁴, Gabriel Kolle¹¹, Felicity Newell⁴, Mark Pinese¹, R. Scott Mead^{1,12}, Jeremy L. Humphris¹, Warren Kaplan¹, Marc D. Jones¹, Emily K. Colvin¹, Adnan M. Nagrial¹, Emily S. Humphrey¹, Angela Chou^{1,12}, Venessa T. Chin¹, Lorraine A. Chantrill¹, Amanda Mawson¹, Jaswinder S. Samra^{9,13}, James G. Kench^{1,10,14}, Jessica A. Lovell¹, Roger J. Daly¹, Neil D. Merrett^{2,15}, Christopher Toon¹, Krishna Epari¹⁶, Nam Q. Nguyen¹⁷, Andrew Barbour¹⁸, Nikolajs Zeps¹⁹, Australian Pancreatic Cancer Genome Initiative⁴, Nipun Kakkar⁵, Fengmei Zhao⁵, Yuan Qing Wu⁵, Min Wang⁵, Donna M. Muzny⁵, William E. Fisher^{6,20}, F. Charles Brunicaudi²¹, Sally E. Hodges^{6,20}, Jeffrey G. Reid⁵, Jennifer Drummond⁵, Kyle Chang⁵, Yi Han⁵, Lora R. Lewis⁵, Huyen Dinh⁵, Christian J. Buhay⁵, Timothy Beck⁷, Lee Timms⁷, Michelle Sam⁷, Kimberly Begley⁷, Andrew Brown⁷, Deepa Pai⁷, Ami Panchal⁷, Nicholas Buchner⁷, Richard De Borja⁷, Robert E. Denroche⁷, Christina K. Yung⁷, Stefano Serra²², Nicole Onetto⁷, Debabrata Mukhopadhyay²³, Ming-Sound Tsao²², Patricia A. Shaw²², Gloria M. Petersen²³, Steven Gallinger^{22,24}, Ralph H. Hruban²⁵, Anirban Maitra²⁵, Christine A. Iacobuzio-Donahue²⁵, Richard D. Schulick²⁶, Christopher L. Wolfgang²⁶, Richard A. Morgan²⁵, Rita T. Lawlor²⁷, Paola Capelli²⁸, Vincenzo Corbo²⁷, Maria Scardoni²⁸, Giampaolo Tortora²⁹, Margaret A. Tempero³⁰, Karen M. Mann³¹, Nancy A. Jenkins³¹, Pedro A. Perez-Mancera³², David J. Adams³³, David A. Largaespa³⁴, Lodewyk F. A. Wessels³⁵, Alistair G. Rust³³, Lincoln D. Stein³⁴, David A. Tuveson³², Neal G. Copeland³¹, Elizabeth A. Musgrove^{1,36}, Aldo Scarpa^{27,28}, James R. Eshleman²⁵, Thomas J. Hudson⁷, Robert L. Sutherland^{1,36†}, David A. Wheeler⁵, John V. Pearson⁴, John D. McPherson⁷, Richard A. Gibbs⁵ & Sean M. Grimmond⁴

¹The Kinghorn Cancer Centre, 370 Victoria Street, Darlinghurst, and the Cancer Research Program, Garvan Institute of Medical Research, 384 Victoria Street, Darlinghurst, Sydney, New South Wales 2010, Australia. ²Department of Surgery, Bankstown Hospital, Eldridge Road, Bankstown, Sydney, New South Wales 2200, Australia. ³South Western Sydney Clinical School, Faculty of Medicine, University of New South Wales, Liverpool, New South Wales 2170, Australia. ⁴Queensland Centre for Medical Genomics, Institute for Molecular Bioscience, The University of Queensland, St Lucia, Brisbane, Queensland 4072, Australia. ⁵Department of Molecular and Human Genetics, Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, MS226, Houston, Texas 77030-3411, USA. ⁶Michael E. DeBakey Department of Surgery, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030-3411, USA. ⁷Ontario Institute for Cancer Research, Toronto M5G 0A3, Canada. ⁸School of Environmental & Life Sciences, University of Newcastle, Ourimbah, New South Wales 2258, Australia. ⁹Department of Anatomical Pathology, Royal North Shore Hospital, St Leonards, Sydney, New South Wales 2065, Australia. ¹⁰University of Sydney, Sydney, New South Wales 2006, Australia. ¹¹Life Technologies, Brisbane, Queensland 4000, Australia. ¹²Department of Anatomical Pathology, Sydnpath, St Vincent's Hospital, Sydney, New South Wales 2010, Australia. ¹³Department of Surgery, Royal North Shore Hospital, St Leonards, Sydney, New South Wales 2065, Australia. ¹⁴Tissue Pathology and Diagnostic Oncology, Royal Prince Alfred Hospital, Camperdown, New South Wales 2050, Australia. ¹⁵School of Medicine, University of Western Sydney, Penrith, New South Wales 2175, Australia. ¹⁶Department of Surgery, Fremantle Hospital, Alma Street, Fremantle, Western Australia 6160, Australia. ¹⁷Department of Gastroenterology, Royal Adelaide Hospital, North Terrace, Adelaide, South Australia 5000, Australia. ¹⁸Department of Surgery, The University of Queensland, Princess Alexandra Hospital, Ipswich Road, Woollongabba, Queensland 4102, Australia. ¹⁹School of Surgery M507, University of Western Australia, 35 Stirling Hwy, Nedlands 6009, Australia and St John of God Pathology, 12 Salvado Road, Subiaco, Western Australia 6008, USA. ²⁰The Elkins Pancreas Center at Baylor College of Medicine, 6620 Main Street, Suite 1450, Houston, Texas 77030, USA. ²¹David Geffen School of Medicine at UCLA, Los Angeles, California 90024, USA. ²²University Health Network, Toronto M5G 2C4, Canada. ²³Mayo Clinic, Rochester, Minnesota 55905, USA. ²⁴Mount Sinai Hospital, Toronto M5G 1X5, Canada. ²⁵Department of Pathology, The Sol Goldman Pancreatic Cancer Research Center, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21231, USA. ²⁶Department of Surgery, The Sol Goldman Pancreatic Cancer Research Center, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21231, USA. ²⁷ARC-NET center for applied research on cancer, University and Hospital Trust of Verona, Verona 37134, Italy. ²⁸Department of Pathology and Diagnostics, University of Verona, Verona 37134, Italy. ²⁹Department of Surgery and Oncology, University and Hospital Trust of Verona, Verona 37134, Italy. ³⁰Division of Hematology and Oncology, University of California, San Francisco 94115, USA. ³¹Cancer Research Program, The Methodist Hospital Research Institute, 6670 Bertner Avenue, Houston, Texas 77030, USA. ³²Li Ka Shing Centre, Cambridge Research Institute, Cancer Research UK, and Department of Oncology, Robinson Way, Cambridge CB2 0RE, UK. ³³Experimental Cancer Genetics, Wellcome Trust Sanger Institute, Hinxton CB10 1HH, UK. ³⁴Masonic Cancer Center, University of Minnesota, Minneapolis, Minnesota 55455, USA. ³⁵Bioinformatics and Statistics, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands. ³⁶St Vincent's Clinical School, Faculty of Medicine, University of New South Wales, New South Wales 2010, Australia.

*Lists of participants and their affiliations appear below.
†Deceased.

Australian Pancreatic Cancer Genome Initiative

The Kinghorn Cancer Centre, Garvan Institute of Medical Research Andrew V. Biankin¹, Amber L. Johns¹, Amanda Mawson¹, David K. Chang¹, Christopher J. Scarlett¹, Mary-Anne L. Brancato¹, Sarah J. Rowe¹, Skye L. Simpson¹, Mona Martyn-Smith¹, Michelle T. Thomas¹, Lorraine A. Chantrill¹, Venessa T. Chin¹, Angela Chou¹, Mark J.

Cowley¹, Jeremy L. Humphris¹, Marc D. Jones¹, R. Scott Mead¹, Adnan M. Nagrial¹, Marina Pajic¹, Jessica Pettit¹, Mark Pinese¹, Ilse Rooman¹, Jianmin Wu¹, Jiang Tao¹, Renee DiPietro¹, Clare Watson¹, Rachel Wong¹, Andreia V. Pinho¹, Marc Giry-Laterriere¹, Roger J. Daly¹, Elizabeth A. Musgrove¹, Robert L. Sutherland¹; **Queensland Center for Medical Genomics, Institute for Molecular Bioscience** Sean M. Grimmond², Nicola Waddell², Karin S. Kassahn², David K. Miller², Peter J. Wilson², Ann-Marie Patch², Sarah Song², Ivon Harliwong², Senel Idrisoglu², Craig Nourse², Ehsan Nourbakhsh², Suzanne Manning², Shivangi Wani², Milena Gongora², Matthew Anderson², Oliver Holmes², Conrad Leonard², Darrin Taylor², Scott Wood², Qinying Xu², Katia Nones², J. Lynn Fink², Angelika Christ², Tim Bruxner², Nicole Cloonan², Felicity Newell², John V. Pearson²; **Royal North Shore Hospital** Jaswinder S. Samra³, Anthony J. Gill³, Nick Pavlakis³, Alex Guminski³, Christopher Toon³; **Bankstown Hospital** Andrew V. Biankin⁴, Ray Asghari⁴, Neil D. Merrett⁴, David K. Chang⁴, Darren A. Pavey⁴, Amitabha Das⁴; **Liverpool Hospital** Peter H. Cosman⁵, Kasim Ismail⁵, Chelsie O'Connor⁵; **Westmead Hospital** Vincent W. Lam⁶, Duncan McLeod⁶, Henry C. Pleass⁶, Arthur Richardson⁶, Virginia James⁶; **Royal Prince Alfred Hospital** James G. Kench⁷, Caroline L. Cooper⁷, David Joseph⁷, Charbel Sandroussi⁷, Michael Crawford⁷, James Gallagher⁷; **Fremantle Hospital** Michael Textler⁸, Cindy Forrest⁸, Andrew Laycock⁸, Krishna P. Epari⁸, Mo Ballal⁸, David R. Fletcher⁸, Sanjay Mukhedkar⁸; **Sir Charles Gairdner Hospital** Nigel A. Spry⁹, Bastiaan DeBoer⁹, Ming Chai⁹; **St John of God Healthcare** Nikolajs Zeps¹⁰, Maria Beilin¹⁰, Kynan Feeney¹⁰; **Royal Adelaide Hospital** Nam Q. Nguyen¹¹, Andrew R. Ruskiewicz¹¹, Chris Worthley¹¹, Chuan P. Tan¹¹, Tamara Debrencini¹¹; **Flinders Medical Centre** John Chen¹², Mark E. Brooke-Smith¹², Virginia Papangelis¹²; **Greenslopes Private Hospital** Henry Tang¹³, Andrew P. Barbour¹³; **Envoi Pathology** Andrew D. Clouston¹⁴, Patrick Martin¹⁴; **Princess Alexandra Hospital** Thomas J. O'Rourke¹⁵, Amy Chiang¹⁵, Jonathan W. Fawcett¹⁵, Kellee Slater¹⁵, Shinn Yeung¹⁵, Michael Hatzifotis¹⁵, Peter Hodgkinson¹⁵; **Austin Hospital** Christopher Christophi¹⁶, Mehrdad Nikfarjam¹⁶, Angela Mountain¹⁶; **Victorian Cancer Biobank**¹⁷; **Johns Hopkins Medical Institutes** James R. Eshleman¹⁸, Ralph H. Hruban¹⁸, Anirban Maitra¹⁸, Christine A. Iacobuzio-Donahue¹⁸, Richard D. Schulick¹⁸, Christopher L. Wolfgang¹⁸, Richard A. Morgan¹⁸, Mary B.

Hodgin¹⁸; **ARC-NET Center for Applied Research on Cancer** Aldo Scarpa¹⁹, Rita T. Lawlor¹⁹, Paola Capelli¹⁹, Stefania Beghelli¹⁹, Vincenzo Corbo¹⁹, Maria Scardoni¹⁹, Paolo Pederzoli¹⁹, Giampaolo Tortora¹⁹, Claudio Bassi¹⁹; **University of California, San Francisco** Margaret A. Tempero²⁰

¹The Kinghorn Cancer Centre, Cancer Research Program, Garvan Institute of Medical Research, 370 Victoria Street, Darlinghurst, Sydney, New South Wales 2010, Australia. ²Queensland Center for Medical Genomics, Institute for Molecular Bioscience, University of Queensland, St Lucia, Queensland 4072, Australia. ³Royal North Shore Hospital, Westbourne Street, St Leonards, New South Wales 2065, Australia. ⁴Bankstown Hospital, Eldridge Road, Bankstown, New South Wales 2200, Australia. ⁵Liverpool Hospital, Elizabeth Street, Liverpool, New South Wales 2170, Australia. ⁶Westmead Hospital, Hawkesbury and Darcy Roads, Westmead, New South Wales 2145, Australia. ⁷Royal Prince Alfred Hospital, Missenden Road, Camperdown, New South Wales 2050, Australia. ⁸Fremantle Hospital, Alma Street, Fremantle, Western Australia 6959, Australia. ⁹Sir Charles Gairdner Hospital, Hospital Avenue, Nedlands, Western Australia 6009, Australia. ¹⁰St John of God Healthcare, 12 Salvado Road, Subiaco, Western Australia 6008, Australia. ¹¹Royal Adelaide Hospital, North Terrace, Adelaide, South Australia 5000, Australia. ¹²Flinders Medical Centre, Flinders Drive, Bedford Park, South Australia 5042, Australia. ¹³Greenslopes Private Hospital, Newdegate Street, Greenslopes, Queensland 4120, Australia. ¹⁴Envoi Pathology, 1/49 Butterfield Street, Herston, Queensland 4006, Australia. ¹⁵Princess Alexandra Hospital, Cornwall Street & Ipswich Road, Woolloongabba, Queensland 4102, Australia. ¹⁶Austin Hospital, 145 Studley Road, Heidelberg, Victoria 3084, Australia. ¹⁷Victorian Cancer Biobank, 1 Rathdowne Street, Carlton, Victoria 3053, Australia. ¹⁸Johns Hopkins Medical Institutes, 600 North Wolfe Street, Baltimore, Maryland 21287, USA. ¹⁹ARC-NET Center for Applied Research on Cancer, University of Verona, Via dell'Artigliere, 19 37129 Verona 37134, Italy. ²⁰University of California, San Francisco, 1600 Divisadero Street, San Francisco, California 94115, USA.

Broad and potent neutralization of HIV-1 by a gp41-specific human antibody

Jinghe Huang^{1*}, Gilad Ofek^{2*}, Leo Laub¹, Mark K. Louder², Nicole A. Doria-Rose², Nancy S. Longo², Hiromi Imamichi¹, Robert T. Bailer², Bimal Chakrabarti³, Shailendra K. Sharma³, S. Munir Alam⁴, Tao Wang², Yongping Yang², Baoshan Zhang², Stephen A. Migueles¹, Richard Wyatt³, Barton F. Haynes⁴, Peter D. Kwong², John R. Mascola² & Mark Connors¹

Characterization of human monoclonal antibodies is providing considerable insight into mechanisms of broad HIV-1 neutralization. Here we report an HIV-1 gp41 membrane-proximal external region (MPER)-specific antibody, named 10E8, which neutralizes ~98% of tested viruses. An analysis of sera from 78 healthy HIV-1-infected donors demonstrated that 27% contained MPER-specific antibodies and 8% contained 10E8-like specificities. In contrast to other neutralizing MPER antibodies, 10E8 did not bind phospholipids, was not autoreactive, and bound cell-surface envelope. The structure of 10E8 in complex with the complete MPER revealed a site of vulnerability comprising a narrow stretch of highly conserved gp41-hydrophobic residues and a critical arginine or lysine just before the transmembrane region. Analysis of resistant HIV-1 variants confirmed the importance of these residues for neutralization. The highly conserved MPER is a target of potent, non-self-reactive neutralizing antibodies, suggesting that HIV-1 vaccines should aim to induce antibodies to this region of HIV-1 envelope glycoprotein.

Induction of an antibody response capable of neutralizing diverse HIV-1 isolates is a critical goal for vaccines that protect against HIV-1 infection. Potentially the greatest obstacle to achieving this goal is the extraordinary diversity that develops in the target of neutralizing antibodies, the envelope glycoprotein (Env). Although vaccines have thus far failed to induce broadly neutralizing antibody responses, there are examples of chronically infected patients with sera that neutralize highly diverse HIV-1 isolates^{1–8}. These individuals provide evidence that it is possible for the human antibody response to neutralize highly diverse strains of HIV-1, although the mechanisms by which such responses are induced or mediated remain incompletely understood^{9,10}.

Recently, isolation and characterization of human monoclonal antibodies from cells of chronically infected patients have provided considerable advances in understanding the specificities and mechanisms underlying broadly neutralizing antibody responses to HIV-1. Env exists on the virion and infected-cell surface as a trimer of heterodimers made up of gp120 and gp41 subunits. For some time, only a small number of broadly neutralizing monoclonal antibodies had been isolated, consisting of one antibody that binds the CD4-binding site on gp120 (b12), one that binds a glycan configuration on the outer domain of gp120 (2G12), and three that bind the membrane-proximal external region (MPER) on gp41 (2F5, Z13e1 and 4E10)^{11–13}. More recently, considerably more broad and potent antibodies have been discovered that target the CD4-binding site of the envelope protein^{14–17} (for which VRC01 is a prototype) and glycan-containing regions of the variable 1 (V1)/V2 and V3 regions of gp120^{4,18–20} (for which PG9 and PGT128 are prototypes). The specificities of these new antibodies are providing important information regarding antigen targets on Env to which the humoral immune response might be directed to mediate broad and potent neutralization. However, evidence for these specificities in many chronically infected patients within our HIV-1-infected cohort¹ is lacking, suggesting that broad and potent neutralization may be mediated by other specificities.

Here we report isolation of a broad and potent gp41 MPER-specific human monoclonal antibody, 10E8, from an HIV-1-infected individual with high neutralization titres. 10E8 is among the most broad and potent antibodies thus far described, and lacks many of the characteristics previously thought to limit the usefulness of MPER-specific antibodies in vaccines or passive therapies, including lipid binding and autoreactivity. In addition, the crystal structure of 10E8, along with biochemical binding studies, demonstrate that the breadth of 10E8 is mediated by its unique mode of recognition of a structurally conserved site of vulnerability within the gp41 MPER.

10E8 isolation and neutralizing properties

To understand the specificities and binding characteristics that underlie a broadly neutralizing antibody response we developed techniques that permitted isolation of human monoclonal antibodies without previous knowledge of specificity²⁰. Serum from one donor, N152, exhibited neutralizing breadth and potency in the top 1% of our cohort against a 20 cross-clade pseudovirus panel (Supplementary Table 1)¹. Peripheral blood CD19⁺IgM[−]IgD[−]IgA[−] memory B cells from this patient were sorted and expanded for 13 days with interleukin (IL)-2, IL-21 and CD40-ligand expressing feeder cells. The supernatants of ~16,500 B-cell cultures were screened and IgG genes from wells with neutralization activity were cloned and re-expressed²¹ and two novel antibodies (10E8 and 7H6) were isolated.

Nucleotide sequence analysis of DNA encoding 10E8 and 7H6 revealed that both were IgG3 antibodies and were somatic variants of the same IgG clone. These antibodies were derived from *IGHV3-15*05* and *IGLV3-19*01* germline genes, and were highly somatically mutated in variable genes of both heavy chain (21%) and λ light chain (14%) compared to germ line. These antibodies also possessed a long heavy-chain complementarity-determining 3 region (CDR H3) loop composed of 22 amino acids (Fig. 1a). The heavy chains of 10E8 and

¹HIV-Specific Immunity Section, Laboratory of Immunoregulation, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA. ²Vaccine Research Center, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA. ³IAVI Neutralizing Antibody Center, The Scripps Research Institute, Department of Immunology and Microbial Sciences, La Jolla, California 92037, USA. ⁴Duke Human Vaccine Institute, Duke University, Durham, North Carolina 27710, USA.

*These authors contributed equally to this work.

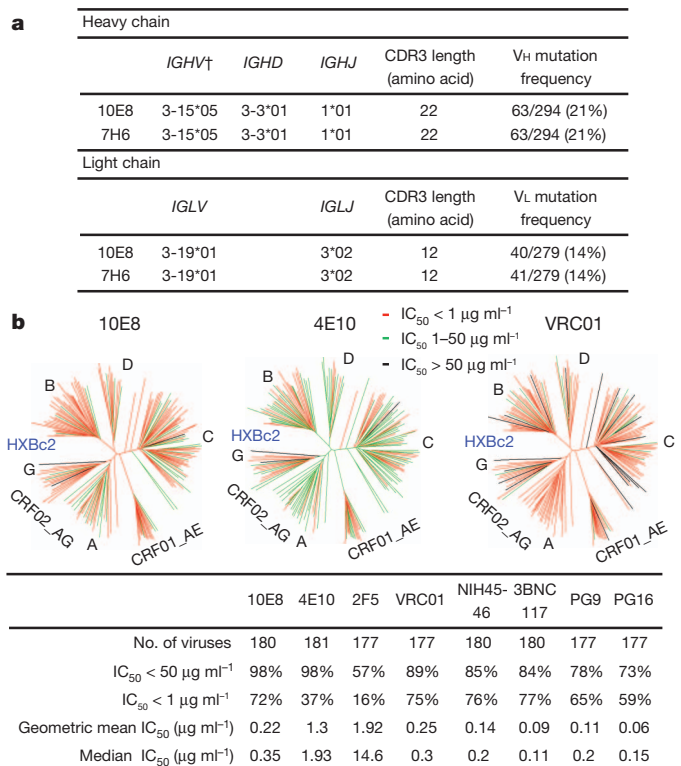


Figure 1 | Analyses of 10E8 sequence and neutralization. **a**, Inferred germline genes encoding the variable regions of 10E8 and 7H6. Dagger footnote symbol indicates that germline alleles were determined using the IMGT database (<http://imgt.org>). **b**, Neutralizing activity of antibodies against a 181-isolate Env-pseudovirus panel. Dendrograms indicate the gp160 protein distance of HIV-1 primary isolate Env glycoproteins. Letters indicate clade. Data below the dendrogram show the number of tested viruses, the percentage of viruses neutralized and the geometric mean IC₅₀ for viruses neutralized with an IC₅₀ < 50 µg ml⁻¹. Median titres are based on all tested viruses, including those with IC₅₀ > 50 µg ml⁻¹, which were assigned a value of 100.

7H6 were identical and there were only two residue differences in the light chain (Supplementary Fig. 1)²².

To assess neutralization activity of the clonal variants, they were initially tested against 5 Env pseudoviruses (Supplementary Table 1a), and monoclonal antibody 10E8 was selected for further study. To determine whether the neutralization activity of 10E8 was representative of the overall neutralization specificity present in patient N152 donor serum, the neutralization panel was expanded to 20 Env pseudoviruses, and 10E8 was tested in parallel with N152 donor serum. Although there were some similarities in the pattern of neutralization of highly resistant variants, a correlation of the neutralization half-maximum inhibitory concentration (IC₅₀) of monoclonal antibody 10E8 and inhibitory dilution (ID₅₀) of N152 serum did not achieve statistical significance ($P = 0.11$; Supplementary Fig. 2 and Supplementary Table 1b). This finding indicates that although 10E8 may have a major role, the full breadth of neutralization of HIV-1 by N152 serum is probably mediated by an amalgam of 10E8-like or other antibodies.

To compare the neutralization potency and breadth of 10E8 with other broadly neutralizing anti-HIV-1 antibodies, 10E8 was then tested in a 181-isolate Env-pseudovirus panel in parallel with 4E10, 2F5, VRC01, NIH45-46, 3BNC117, PG9 and PG16 (Fig. 1b and Supplementary Table 1c). At an IC₅₀ below 50 µg ml⁻¹, 10E8 neutralized 98% of the tested pseudoviruses compared to 98% for 4E10 and 89% for VRC01. However, at an IC₅₀ below 1 µg ml⁻¹, 10E8 neutralized 72% of the tested viruses compared to 37% for 4E10. The median and geometric mean IC₅₀ values for 10E8 were below 1 µg ml⁻¹. Thus, 10E8 mediates broad and potent neutralization against a large range

of viruses and the potency is comparable to some of the best available monoclonal antibodies.

10E8 epitope specificity and binding

To map the epitope of the 10E8 antibody, we tested binding to different subregions of Env by enzyme-linked immunosorbent assay (ELISA). 10E8 bound strongly to gp140, gp41 and the 4E10-specific MPER peptide, but not to gp120 (Fig. 2a). To map further the 10E8 epitope within the MPER, we examined binding of 10E8 to overlapping peptides corresponding to the 2F5 (656–671), Z13e1 (666–677) and 4E10 (671–683) specificities. 10E8 bound to the full MPER and the 4E10-specific peptides, but not 2F5- or Z13e1-specific peptides. Within the 4E10 epitope, when a peptide with a truncated carboxy terminus was tested, 4E10.19 (671–680), 10E8 binding was weakened considerably, indicating that the three terminal amino acids of the MPER—Tyr681, Ile682 and Arg683—were crucial for 10E8 binding (Supplementary Fig. 3a). Consistent with these results, only the full MPER and 4E10-specific peptides blocked 10E8-mediated neutralization of the chimeric C1 virus, which contains the HIV-2 Env with the HIV-1 MPER (Supplementary Fig. 3b). Taken together, these data indicate that the minimal 10E8 epitope is located within residues 671–683 of the MPER, although additional contacts towards the amino terminus of the MPER could not be excluded.

To map more precisely the epitope of 10E8, a panel of alanine mutant peptides scanning MPER residues 671–683 was used to block 10E8 neutralization of the chimeric C1 virus (Fig. 2b)²³. MPER peptides with alanine substitutions at Trp 672, Phe 673 or Thr 676 failed to block 4E10 or 10E8 neutralization, indicating that these residues are critical for both 4E10 and 10E8 binding. Residues Asn 671 and Arg 683, both of which are not required for 4E10 binding, were found to be critical for 10E8 binding and neutralization (Supplementary Table 2 and Fig. 2b). We also tested the ability of 10E8 to neutralize HIV-1 JR2 pseudoviruses with alanine substitutions in MPER residues 660–683 (Supplementary Table 3). Consistent with the effects of alanine substitutions on peptide binding, residues Asn 671 and Arg 683 were critical for 10E8, but not 4E10, neutralization. Individual alanine substitutions at residues 671–673, 680 and 683 resulted in reduced neutralization sensitivity to 10E8 most apparent at the IC₉₀ level rather

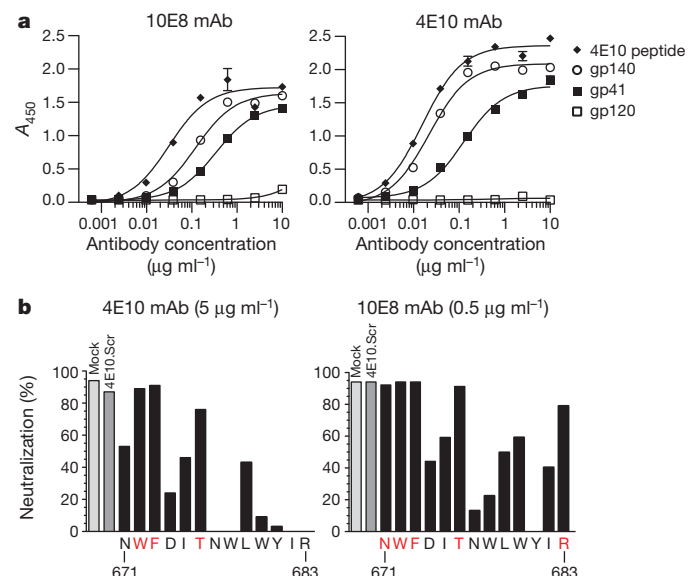


Figure 2 | Binding specificity of 10E8. **a**, ELISA binding of monoclonal antibody (mAb) 10E8 or 4E10 to gp140, gp120, gp41 or 4E10 peptide. Error bars denote one standard error of the mean (s.e.m.). **b**, Inhibition of monoclonal antibody 10E8 or 4E10 neutralization of C1 HIV-2/HIV-1 MPER virus by 4E10 alanine scanning peptides. Residues shown in red indicate positions for which the alanine mutant peptide did not block neutralization.

than the IC₅₀ level. Although the mechanism for this phenomenon is unclear, a similar effect has been observed previously when MPER mutations cause partial resistance to 4E10²⁴. Taken together, these results indicated that 10E8 recognized a novel epitope that overlaps the known 4E10 and Z13e1 epitopes but differs in a critical dependence on binding to Asn 671 and Arg 683, the last residue of the MPER.

We next investigated whether the greater neutralization potency of 10E8 compared to other MPER antibodies was a result of higher binding affinity to the MPER. Capture of a biotinylated peptide encompassing the full MPER (656–683) to a surface-plasmon resonance (SPR) chip allowed the binding kinetics of antigen-binding fragments (Fabs) 10E8, 2F5 and 4E10 to be examined. In contrast to its higher neutralization potency, the dissociation constant (K_d) of 10E8 to this MPER peptide was weaker than that of 2F5 and 4E10: 17 nM for 10E8 versus 3.8 nM for 2F5 and 0.74 nM for 4E10 (Supplementary Fig. 4). Therefore, the affinity of 10E8 for the MPER in a soluble peptide format did not explain its greater neutralization potency compared to other MPER-specific antibodies.

Prevalence of 10E8-like antibodies

We next investigated the prevalence of MPER-specific and 10E8-like neutralizing antibodies in our cohort of HIV-1-infected donors. We selected 78 sera from our cohort with a neutralization ID₅₀ > 100 against at least one pseudovirus in a five-virus mini-panel¹. The median time since diagnosis of these donors was 13.5 years, median CD4 count was 557 cells μl^{-1} , median plasma HIV RNA 5,573 copies ml^{-1} , and they were not receiving antiretrovirals. We tested neutralization against the HIV-2/HIV-1 chimaera C1 (Supplementary Table 4)²⁵. Of 78 sera, 21 exhibited neutralization activity against the HIV-2/HIV-1 C1 virus (Supplementary Table 5). To map the region that was targeted by these sera, neutralization was measured using seven HIV-2/HIV-1 chimaeras containing subdomains of the MPER (Supplementary Table 4)²⁵. Of the 21 sera with neutralization activity against the entire MPER, 8 exhibited a neutralization pattern similar to that observed for 10E8, which entailed neutralization of only those HIV-2/HIV-1 chimaeric viruses that contained the terminal residue of the MPER Arg 683 (C4, C4GW and C8; Supplementary Table 5). For further confirmation of these results, we used peptides corresponding to different portions of the MPER to block sera neutralization of the HIV-2/HIV-1 chimaera C1 (Supplementary Table 6). Of the eight sera found to have a 10E8-like pattern based on neutralization of the chimaeras, three were blocked by peptides consistent with 10E8-like activity. An additional three of the eight 10E8-like sera were blocked by peptides in a pattern consistent with a combination of 10E8- and Z13e1-like antibodies. The six patients whose sera had 10E8-like activity did not differ from the remaining 72 patients with regard to clinical course or HIV neutralization (Supplementary Fig. 5, legend). Overall, 27% of the tested patient sera exhibited anti-MPER neutralizing activity. This prevalence is considerably higher than observed in previous work, possibly related to selection of donors with known neutralizing activity^{8,26–28}. Furthermore, 8% of the tested sera had 10E8-like antibodies (Supplementary Fig. 5), indicating that 10E8-like antibodies are not rare.

Analysis of 10E8 autoreactivity

A property common to the previously characterized MPER monoclonal antibodies 2F5 and 4E10 is that they crossreact with self-antigens²⁹. In addition, binding to both the cell membrane and the Env trimer is thought to be important for optimal neutralization by these antibodies and this autoreactivity may be an obstacle to the elicitation of similar antibodies by a vaccine^{29,30}. Surface plasmon resonance analysis showed that 10E8 did not bind to anionic phospholipids, such as phosphatidyl choline-cardiolipin (PC-CLP) and phosphatidyl choline-phosphatidyl serine (PC-PS) liposomes (Fig. 3a). 10E8 also did not bind HEp-2 epithelial cells, in contrast to 2F5 and 4E10 that bound in a cytoplasmic and nuclear pattern (Fig. 3b). Additionally, 10E8 did not bind auto-antigens, such as Sjogren's syndrome antigens A and B, Smith antigen,

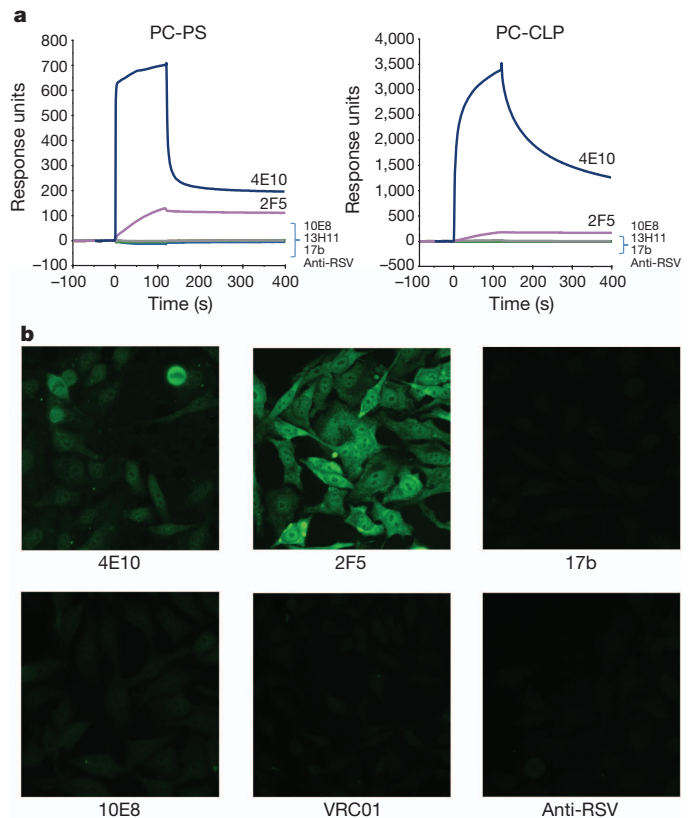


Figure 3 | Analysis of 10E8 autoreactivity. **a**, SPR analysis of 10E8 binding to anionic phospholipids. 10E8 was injected over PC-CLP liposomes or PC-PS liposome immobilized on the Biacore L1 sensor chip. 4E10 and 2F5 were used as positive controls and 13H11, 17b and anti-RSV F protein as negative controls. **b**, Reactivity of 10E8 with HEp-2 epithelial cells. Controls are as above with VRC01 added as an additional negative control. Antibody concentration was 25 $\mu\text{g ml}^{-1}$. All pictures are shown at $\times 400$ magnification.

ribonucleoprotein, scleroderma 70 antigen, Jo1 antigen, centromere B and histone (Supplementary Table 7). Taken together, these results suggest that 10E8, in contrast to other MPER antibodies, is not autoreactive.

Virion accessibility of 10E8

The 2F5 and 4E10 antibodies have been shown to bind relatively poorly to the HIV-1 envelope spike on the surface of infected cells or to cell-free virions, and react more efficiently after Env engagement of the CD4 receptor³¹. We measured binding to cleaved, full-length envelope spikes on HIV_{JRFL} transfected cells (Supplementary Fig. 6a). Although 10E8 bound less efficiently than other antibodies such as VRC01 or F105, where accessibility is not an issue, it bound more efficiently than either 2F5 or 4E10. In contrast to results of alanine substitution, a mutation in the 4E10 (F673S) region in full-length HIV_{JRFL} envelope spikes enhanced 10E8 binding although the mechanism remains unclear. A mutation in the 2F5 (K665E) region had no influence on 10E8 binding. These data indicate that 10E8 has modestly greater access to the MPER epitope on the cell surface than either 2F5 or 4E10.

To assess binding to cell-free virus, we incubated virions with antibody, washed out unbound antibody, and tested neutralization^{31–33}. During washing, antibodies that cannot access their Env target on free virions will be largely removed and therefore neutralization will be diminished. As a control, neutralization of the HXBc2 isolate was not diminished by washing, because the MPER region is accessible on this laboratory-adapted isolate³¹. Washing also had little impact on neutralization of JRFL by VRC01. Consistent with previous work, 2F5 and 4E10 neutralization of most virus isolates tested was substantially diminished after washing (Supplementary Fig. 6b)^{31,34}. In contrast to 2F5 and

4E10, washing had a smaller effect on 10E8 neutralization of most viruses tested, as measured by the area under the curve or analysis of the fold-change in neutralization at a fixed inhibitory concentration (Supplementary Fig. 6c). Although 10E8 is not fully able to access its epitope on the native viral spike similarly to VRC01, under most experimental conditions tested it was better able to access its epitope than either 2F5 or 4E10.

Structure of the 10E8–gp41 complex

To provide an atomic-level understanding of the interaction of 10E8 with HIV-1, we crystallized the Fab of 10E8 in complex with a peptide encompassing the entire 28-residue gp41 MPER (residues 656–683). Monoclinic crystals diffracted to 2.1 Å resolution, and structure solution and refinement to $R_{\text{cryst}} = 18.01\%$ ($R_{\text{free}} = 21.76\%$) revealed two complexes in the asymmetric unit (heretofore referred to as complexes 1 and 2) (Supplementary Table 8). Overall, 10E8 bound to one face of the MPER peptide, which formed two helices, each 15–20 Å in length and oriented 100° relative to each other (Fig. 4a). Electron density was observed for the entire MPER, ranging from Asn 656 to Arg 683 (Leu 660 to Arg 683 for complex 2), with the highest degree of ordered density observed from residue Trp 666 within the N-terminal helix through to Arg 683 of the C-terminal helix (Supplementary Fig. 7). Analysis of main-chain dihedral angles (Supplementary Table 9) indicated that the N-terminal α -helix extends from residue Asn 657 to Ala 667, tightens into a 3_{10} -helix

between residues Ser 668 and Leu 669, before turning at residues Trp 670 and Asn 671. The C-terminal α -helix, capped by Asn 671, starts at residue Trp 672 and extends to residue Arg 683, the final residue of the MPER (Fig. 4a, b).

The 10E8 antibody contacts the gp41 MPER primarily through its heavy chain, although crucial contacts are also mediated by the light chain CDR L3 (Fig. 4a–c and Supplementary Tables 10–12). Three predominant loci of interaction are observed between the antibody and gp41 (Supplementary Tables 13 and 14): one between residues of the tip of the CDR H3 loop and the tip of the C-terminal helix of the peptide, a second between residues of the CDR H2 loop and residues of the hinge region of the peptide, and a third at the juncture of the three heavy-chain CDR loops and the light chain CDR L3, which form a hydrophobic cleft that holds residues of the beginning of the MPER C-terminal helix (Fig. 4b).

10E8–gp41 interface

To complement the results observed for the mutagenesis of the highly conserved 10E8 epitope (Fig. 4d, f, h and Supplementary Tables 2 and 3), each residue of the 10E8 paratope, as determined from the crystal structure, was individually mutated to alanine and the resulting 25 10E8 variants assessed for affinity to a soluble MPER peptide. Overall, the most pronounced effects of the alanine mutations on the binding affinity of 10E8 to a soluble MPER peptide occurred within residues of the CDR H3 loop, although mutations

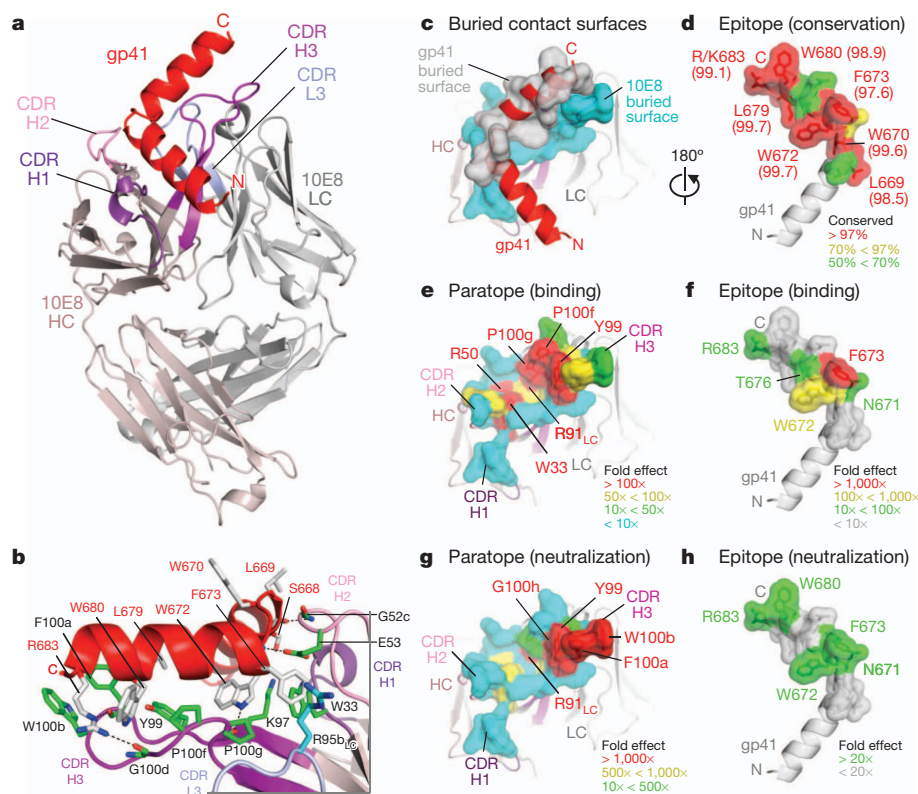


Figure 4 | Crystal structure of 10E8 Fab in complex with its gp41 MPER epitope. **a**, 10E8 recognizes a highly conserved gp41 helix to neutralize HIV-1. Fab 10E8 is shown in ribbon representation (shades of violet for heavy chain (HC) and of grey for light chain (LC)) in complex with a gp41 peptide (red) that encompasses the MPER (Asn 656–Arg 683). **b**, Interface between 10E8 and gp41 with select 10E8 side chains (green, heavy chain; cyan, light chain) and gp41 side chains (grey) in stick representation. In analogy to a hand, the hinge can be viewed as being gripped by a thumb (represented by the CDR H2), the C-terminal helix as being suspended along a corresponding extended forefinger (represented by the CDR H3), and residues that commence the C-terminal helix as being caught in the cleft between the thumb and forefinger (represented

by the juncture of the CDR loops). (See ref. 22 for numbering.) **c**, **d**, Buried contact surfaces and epitope conservation. An examination of the buried contact surface on gp41 (grey; **c**) reveals that highly conserved epitope residues (labelled, **d**) are buried by 10E8 (Supplementary Tables 11 and 12; conservation percentages provided in parentheses). **e**–**h**, Alanine mutagenesis of paratope and epitope. Residues at the tip of the 10E8 CDR H3 loop and within the hydrophobic cleft are crucial for recognition of gp41 and for virus neutralization (Supplementary Tables 15 and 16), as mapped onto the buried 10E8 contact surface (**c**, **e**, **g**). These results mirror the effects of alanine scan mutations of the 10E8 epitope (Supplementary Tables 2 and 3), as mapped onto the buried gp41 contact surface (**f**, **h**).

within the hydrophobic cleft also showed substantial effects (Fig. 4e, Supplementary Table 15 and Supplementary Fig. 8). 10E8 residues identified by an alanine scan as critical for the interaction with gp41 stretched from the cleft all the way to the tip of the CDR H3 (Fig. 4e) and were mirrored by a corresponding stretch of gp41 residues that substantially affected 10E8 binding when mutated to alanine (Fig. 4f).

The same panel of 10E8 alanine mutations was tested for neutralization potency against a panel of five Env pseudoviruses that included both tier 1 and tier 2 viruses (Supplementary Table 16). Similar to the binding data, residues of the 10E8 CDR H3 had marked effects on neutralization, as did residues of the hydrophobic cleft (Fig. 4g). Generally, K_d values of paratope mutants correlated with neutralization (Supplementary Fig. 9). Backbone interactions (on both 10E8 and gp41) also contribute to the interface, especially between the CDR H2 of 10E8 and the hinge region of the MPER, although these are silent in alanine scan analyses. Overall, 10E8 uses a narrow band of residues ($\sim 20 \times 5$ Å) that stretches from the CDR H1 and H2 and extends along most of the CDR H3 to recognize a string of highly conserved hydrophobic gp41 residues, along with a critical charged residue (Arg/Lys 683), that occurs just before the transmembrane region (Fig. 4f, h).

A conserved gp41-neutralization determinant

Several structures of neutralizing antibodies in complex with the MPER of gp41 have been reported previously, including those for antibodies 2F5, Z13e1 and 4E10^{35–39} (Supplementary Fig. 10a). The MPER adopts divergent loop conformations when bound by 2F5 and Z13e1 and an α -helix when bound by 4E10. Comparison of 2F5, Z13e1 and 4E10 epitopes with 10E8-bound gp41 revealed that only the 4E10 epitope has similar secondary structure, with superposition yielding a root mean squared deviation (r.m.s.d.) of 2.49 Å for all atoms of residues 671–683 and 0.98 Å for main-chain atoms (Supplementary Fig. 10b and Supplementary Table 17).

To further compare the recognition of gp41 by 10E8 and 4E10, we examined their angles of epitope approach. As shown in Supplementary Fig. 10c–f, alignment of the recognized MPER helix places 10E8 and 4E10 into similar overall spatial positions. The relative orientations of the recognized helix and the heavy and light chains of the two antibodies, however, differ markedly. With 10E8, the C-terminal helix is perpendicular to the plane bisecting the heavy and light chains (Supplementary Fig. 10c, e); with 4E10, the recognized helix is at the interface between the heavy and light chains (Supplementary Fig. 10d, f). Perhaps relevant to this, 10E8 uses CDR loops almost exclusively in its recognition of gp41, whereas 4E10 incorporates substantial β -strand interactions with gp41.

The differing modes of 10E8 and 4E10 recognition of the conserved C-terminal MPER helix result in a substantial difference in the proportion of the recognized helical face: 10E8 contacts roughly one-third of the helical face, whereas 4E10 contacts over half (Supplementary Fig. 10g, h and Supplementary Tables 18 and 19). The smaller contact surface of 10E8 may provide an explanation for the reduced recognition of lipid surfaces by 10E8 versus 4E10, providing a potential structure-based explanation for reduced autoreactivity of 10E8.

Sequence variation and 10E8 neutralization

To place the specificity and structural data into the context of known variation of the MPER, we analysed viral sequences with resistance to neutralization by 10E8 (Fig. 5a). Of the 181 viruses tested, only 3 were highly resistant to 10E8 with $IC_{50} > 50 \mu\text{g ml}^{-1}$. Each of these viruses had substitutions at positions found to affect neutralization by alanine scanning (Asn 671, Trp 672, Phe 673 and Trp 680). Plasma virus from patient N152, from whom 10E8 was cloned, is also probably resistant to 10E8-mediated neutralization⁴⁰. Sequence analysis of plasma viral RNA revealed rare substitutions at positions Trp 680 and Lys/Arg 683 (Fig. 5a). These residues are typically highly conserved with variation

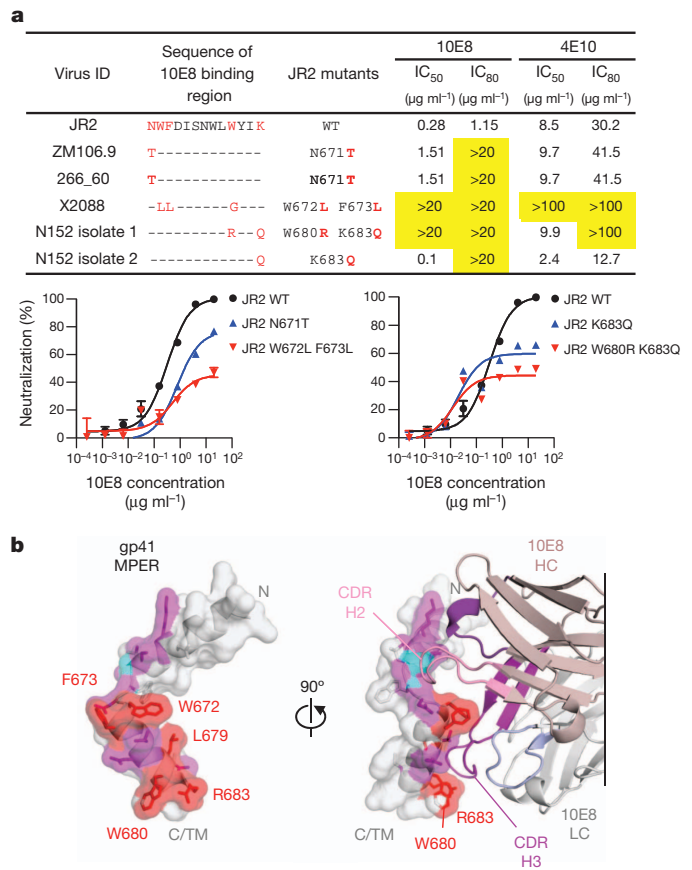


Figure 5 | A site of gp41 vulnerability. **a**, Impact of sequence variation on 10E8 neutralization. Predicted amino acid sequences within the binding epitope of 10E8 for three 10E8-resistant viruses and the patient virus are shown. The 10E8 epitope and differences in sequence compared to the JR2 virus are labelled in red. IC_{50} and IC_{80} values that are >20-fold compared with JR2 wild-type pseudovirus are highlighted in yellow. Error bars denote one s.e.m. **b**, Structural definition of a highly conserved region of gp41 recognized by neutralizing antibodies. Atoms of highly conserved residues that make direct contacts with 10E8 (crystal complex 1) are coloured red and shown in stick representation. Remaining atoms buried by 10E8 are coloured purple, and those making main-chain contacts are coloured cyan. Semi-transparent surfaces of the gp41 MPER are coloured according to the underlying atoms. 90° views are shown, with bound antibody 10E8 in the right panel. The 10E8 CDR H3 interacts with highly conserved hydrophobic residues, whereas the CDR H2 contacts main-chain atoms at the juncture between the N- and C-terminal helices. Many of the unbound residues of the MPER (grey) are hydrophobic, especially those within the C-terminal helix. In the structure of the late fusion intermediate (Supplementary Fig. 11), gp41 residues that contact 10E8 largely face the outside of the helical coiled-coil.

only occurring in 1.17% of 3,730 HIV Env sequences in the Los Alamos Database (<http://www.hiv.lanl.gov>). When the substitutions for the three resistant viruses and the patient viruses were placed on the background of the sensitive JR2 virus, substitutions at Asn671Thr, Trp672Leu and Phe673Leu had a modest effect on the IC_{50} values but raised the IC_{80} values above $20 \mu\text{g ml}^{-1}$. In the structural analysis, direct contacts with 10E8 were not observed at position 671, indicating that the effects on neutralization of Thr or Ala substitutions at this position are mediated by conformational or other effects within gp41. The combination of Trp672Leu and Phe673Leu conferred high-level resistance at the IC_{50} and IC_{80} level. Changes corresponding to the patient's dominant circulating virus had a similar effect. Although Lys/Arg683Gln alone conferred resistance at the IC_{80} level, together Trp680Arg and Lys/Arg683Gln resulted in greater resistance to 10E8 (Fig. 5a). When taken together with the analysis of the 10E8 paratope, these data suggest that in addition to Trp 672, Phe 673 and Trp 680

found in the 4E10 epitope, the additional 10E8-bound residue Lys/Arg683 is critical to neutralization. In addition to other differences in binding based on structural analyses noted above, it is possible that the additional potency of 10E8 compared to 4E10 against naturally occurring viral variants may be mediated through binding of highly conserved residues Trp 680 and Lys/Arg 683 that directly interact with the 10E8 CDRH3.

Discussion

10E8 is a broad and potent neutralizing antibody with important implications for efforts to induce such antibodies with vaccines. Previous MPER antibodies were limited in potency, and had a more limited ability to access MPER on Env of primary isolates. In addition, lipid binding and autoreactivity were thought to be characteristics of MPER antibodies and important obstacles to their elicitation by vaccines^{9,29,30}. However, 10E8 lacks each of these characteristics. In addition, antibodies with a similar specificity were not rare in our chronically infected cohort. This suggests that 10E8-like antibodies were not deleted from the repertoire because of autoreactivity. These results further suggest that 10E8-like antibodies might be raised in a larger fraction of HIV-uninfected persons receiving a vaccine designed to elicit these antibodies without the B-cell defects of chronic HIV infection. Design of such a vaccine will probably require not only presentation of an intact 10E8 epitope but also use of a platform sufficiently immunogenic to drive the evolution of 10E8-like antibodies.

The extraordinary breadth and potency of 10E8 seems to be mediated by its ability to bind highly conserved residues within MPER. Although the epitope of 10E8 overlaps those of known monoclonal antibodies such as 4E10, it differs in recognition surface, angle of approach, lipid binding and self-reactivity. Alanine scanning, structural analysis and paratope analysis each indicate that 10E8 makes crucial contacts with highly conserved residues Trp 672, Phe 673, Trp 676 and Lys/Arg 683. The extraordinary breadth of some potent monoclonal antibodies, for example that bind the CD4 binding site, is thought to be conferred by blocking a functionally important site that is critical for viral entry. Whether 10E8 impairs Env function or simply acts by binding highly conserved residues remains to be determined. Nonetheless, the breadth and potency of 10E8 demonstrates a conserved site of gp41 vulnerability (Fig. 5b) that is an important target antigen for HIV neutralization and that will probably reinvigorate interest in MPER-based HIV vaccine design.

METHODS SUMMARY

Peripheral blood CD19⁺IgM⁺IgD⁺IgA⁺ B cells were sorted by flow cytometry, plated at 4 cells per well, and expanded with cytokines and feeder cells. B-cell culture supernatants were screened by microneutralization against HIV_{MN.03} and HIV_{Bal.26} pseudoviruses. IgG genes from wells with neutralization activity were cloned and re-expressed in 293T cells. Breadth of neutralizing activity was confirmed against a 181-isolate Env pseudovirus panel. Specificity was determined by alanine scanning peptides and mutant Env pseudoviruses. Lipid binding and autoreactivity of 10E8 were measured by SPR plasmon resonance, indirect immunofluorescence on HEP-2 cells and bead arrays. Binding of HIV envelopes on transfected 293 cells was detected by flow cytometry. After pre-incubation with antibody, the impact of washing virions before infecting TZM-bl cells was used to measure access to viral MPER. The frequency of HIV-1⁺ sera with a given specificity was measured by the ability to neutralize HIV-2/HIV-1 chimaeras containing portions of the MPER. 10E8 was co-crystallized with a peptide encompassing the entire 28-residue gp41 MPER (residues 656–683). Structure determination revealed two complexes in the crystal asymmetric unit. Analysis of differences between the two complexes enabled essential interactions to be discerned. The paratope, as defined by residues in the antibody that showed reduced solvent accessibility when in complex with gp41, was subjected to comprehensive alanine scan, with each of the 25 10E8 alanine mutants assessed by SPR for recognition of gp41 and for neutralization of a panel of 5 pseudotyped viruses. The sequence of the patient plasma viral RNA was derived using limiting dilution RT-PCR.

Full Methods and any associated references are available in the online version of the paper.

Received 5 June; accepted 10 September 2012.

Published online 18 September 2012.

1. Doria-Rose, N. A. *et al.* Breadth of human immunodeficiency virus-specific neutralizing activity in sera: clustering analysis and association with clinical variables. *J. Virol.* **84**, 1631–1636 (2010).
2. Stamatatos, L., Morris, L., Burton, D. R. & Mascola, J. R. Neutralizing antibodies generated during natural HIV-1 infection: good news for an HIV-1 vaccine? *Nature Med.* **15**, 866–870 (2009).
3. Sather, D. N. *et al.* Factors associated with the development of cross-reactive neutralizing antibodies during human immunodeficiency virus type 1 infection. *J. Virol.* **83**, 757–769 (2009).
4. Walker, L. M. *et al.* A limited number of antibody specificities mediate broad and potent serum neutralization in selected HIV-1 infected individuals. *PLoS Pathog.* **6**, e1001028 (2010).
5. Simek, M. D. *et al.* Human immunodeficiency virus type 1 elite neutralizers: individuals with broad and potent neutralizing activity identified by using a high-throughput neutralization assay together with an analytical selection algorithm. *J. Virol.* **83**, 7337–7348 (2009).
6. Binley, J. Specificities of broadly neutralizing anti-HIV-1 sera. *Curr. Opin. HIV AIDS* **4**, 364–372 (2009).
7. Moore, P. L. *et al.* Potent and broad neutralization of HIV-1 subtype C by plasma antibodies targeting a quaternary epitope including residues in the V2 loop. *J. Virol.* **85**, 3128–3141 (2011).
8. Gray, E. S. *et al.* Antibody specificities associated with neutralization breadth in plasma from human immunodeficiency virus type 1 subtype C-infected blood donors. *J. Virol.* **83**, 8925–8937 (2009).
9. Haynes, B. F., Kelsoe, G., Harrison, S. C. & Kepler, T. B. B-cell-lineage immunogen design in vaccine development with HIV-1 as a case study. *Nature Biotechnol.* **30**, 423–433 (2012).
10. Walker, L. M. & Burton, D. R. Rational antibody-based HIV-1 vaccine design: current approaches and future directions. *Curr. Opin. Immunol.* **22**, 358–366 (2010).
11. Zwick, M. B. *et al.* Broadly neutralizing antibodies targeted to the membrane-proximal region of human immunodeficiency virus type 1 glycoprotein gp41. *J. Virol.* **75**, 10892–10905 (2001).
12. Burton, D. R. *et al.* Efficient neutralization of primary isolates of HIV-1 by a recombinant human monoclonal antibody. *Science* **266**, 1024–1027 (1994).
13. Muster, T. *et al.* A conserved neutralizing epitope on gp41 of human immunodeficiency virus type 1. *J. Virol.* **67**, 6642–6647 (1993).
14. Bonsignori, M. *et al.* Two distinct broadly neutralizing antibody specificities of different clonal lineages in a single HIV-1-infected donor: implications for vaccine design. *J. Virol.* **86**, 4688–4692 (2012).
15. Wu, X. *et al.* Focused evolution of HIV-1 neutralizing antibodies revealed by structures and deep sequencing. *Science* **333**, 1593–1602 (2011).
16. Scheid, J. F. *et al.* Sequence and structural convergence of broad and potent HIV antibodies that mimic CD4 binding. *Science* **333**, 1633–1637 (2011).
17. Wu, X. *et al.* Rational design of envelope identifies broadly neutralizing human monoclonal antibodies to HIV-1. *Science* **329**, 856–861 (2010).
18. Walker, L. M. *et al.* Broad neutralization coverage of HIV by multiple highly potent antibodies. *Nature* **477**, 466–470 (2011).
19. Bonsignori, M. *et al.* Analysis of a clonal lineage of HIV-1 envelope V2/V3 conformational epitope-specific broadly neutralizing antibodies and their inferred unmutated common ancestors. *J. Virol.* **85**, 9998–10009 (2011).
20. Walker, L. M. *et al.* Broad and potent neutralizing antibodies from an African donor reveal a new HIV-1 vaccine target. *Science* **326**, 285–289 (2009).
21. Tiller, T. *et al.* Efficient generation of monoclonal antibodies from single human B cells by single cell RT-PCR and expression vector cloning. *J. Immunol. Methods* **329**, 112–124 (2008).
22. Kabat, E. A., Foeller, C., Gottesman, K. S., Pery, H. M. & Wu, T. T. *Sequences of Proteins of Immunological Interest: Tabulation and Analysis of Amino Acid and Nucleic Acid Sequences of Precursors, V-regions, C-regions, J-chain, T-cell Receptors for Antigen T-cell Surface Antigens, β 2-microglobulins, Major Histocompatibility Antigens, Thy-1, Complement C-reactive Protein, Thymopoietin, Integrins, Post- γ globulin, α 2-macroglobulins, and Other Related Proteins* 5th edn (US Dept. Health and Human Services, Public Health Service, National Institutes of Health, 1991).
23. Brunel, F. M. *et al.* Structure-function analysis of the epitope for 4E10, a broadly neutralizing human immunodeficiency virus type 1 antibody. *J. Virol.* **80**, 1680–1687 (2006).
24. Zwick, M. B. *et al.* Anti-human immunodeficiency virus type 1 (HIV-1) antibodies 2F5 and 4E10 require surprisingly few crucial residues in the membrane-proximal external region of glycoprotein gp41 to neutralize HIV-1. *J. Virol.* **79**, 1252–1261 (2005).
25. Gray, E. S. *et al.* Neutralizing antibody responses in acute human immunodeficiency virus type 1 subtype C infection. *J. Virol.* **81**, 6187–6196 (2007).
26. Tomaras, G. D. *et al.* Polyclonal B cell responses to conserved neutralization epitopes in a subset of HIV-1-infected individuals. *J. Virol.* **85**, 11502–11519 (2011).
27. Morris, L. *et al.* Isolation of a human anti-HIV gp41 membrane proximal region neutralizing antibody by antigen-specific single B cell sorting. *PLoS ONE* **6**, e23532 (2011).
28. Gray, E. S. *et al.* Broad neutralization of human immunodeficiency virus type 1 mediated by plasma antibodies against the gp41 membrane proximal external region. *J. Virol.* **83**, 11265–11274 (2009).

29. Haynes, B. F. *et al.* Cardioliipin polyspecific autoreactivity in two broadly neutralizing HIV-1 antibodies. *Science* **308**, 1906–1908 (2005).
30. Alam, S. M. *et al.* Role of HIV membrane in neutralization by two broadly neutralizing antibodies. *Proc. Natl Acad. Sci. USA* **106**, 20234–20239 (2009).
31. Chakrabarti, B. K. *et al.* Direct antibody access to the HIV-1 membrane-proximal external region positively correlates with neutralization sensitivity. *J. Virol.* **85**, 8217–8226 (2011).
32. Frey, G. *et al.* A fusion-intermediate state of HIV-1 gp41 targeted by broadly neutralizing antibodies. *Proc. Natl Acad. Sci. USA* **105**, 3739–3744 (2008).
33. Rathinakumar, R., Dutta, M., Zhu, P., Johnson, W. E. & Roux, K. H. Binding of anti-membrane-proximal gp41 monoclonal antibodies to CD4-liganded and -unliganded human immunodeficiency virus type 1 and simian immunodeficiency virus virions. *J. Virol.* **86**, 1820–1831 (2012).
34. Ruprecht, C. R. *et al.* MPER-specific antibodies induce gp120 shedding and irreversibly neutralize HIV-1. *J. Exp. Med.* **208**, 439–454 (2011).
35. Julien, J. P., Bryson, S., Nieva, J. L. & Pai, E. F. Structural details of HIV-1 recognition by the broadly neutralizing monoclonal antibody 2F5: epitope conformation, antigen-recognition loop mobility, and anion-binding site. *J. Mol. Biol.* **384**, 377–392 (2008).
36. Cardoso, R. M. *et al.* Structural basis of enhanced binding of extended and helically constrained peptide epitopes of the broadly neutralizing HIV-1 antibody 4E10. *J. Mol. Biol.* **365**, 1533–1544 (2007).
37. Cardoso, R. M. F. *et al.* Broadly neutralizing anti-HIV antibody 4E10 recognizes a helical conformation of a highly conserved fusion-associated motif in gp41. *Immunity* **22**, 163–173 (2005).
38. Ofek, G. *et al.* Structure and mechanistic analysis of the anti-human immunodeficiency virus type 1 antibody 2F5 in complex with its gp41 epitope. *J. Virol.* **78**, 10724–10737 (2004).
39. Pejchal, R. *et al.* A conformational switch in human immunodeficiency virus gp41 revealed by the structures of overlapping epitopes recognized by neutralizing antibodies. *J. Virol.* **83**, 8451–8462 (2009).
40. Wu, X. *et al.* Selection pressure on HIV-1 envelope by broadly neutralizing antibodies to the conserved CD4-binding site. *J. Virol.* **86**, 5844–5856 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank C. W. Hallahan for statistical analyses. We thank K. Lloyd, R. Parks, J. Eudailey and J. Blinn for performing autoantibody assays. We also thank M. Zwick for providing us with the HIV-1 JR2 MPER alanine mutant pseudovirus plasmids. HIV-2/HIV-1 chimaeras were provided by G. Shaw and L. Morris. We thank J. Stuckey for assistance with figures, and members of the Structural Biology Section and Structural Bioinformatics Core, Vaccine Research Center, for discussions and comments on the manuscript. This project has been funded in part with federal funds from the Intramural Research Programs of NIAID and the National Cancer Institute, National Institutes of Health, under Contract no. HHSN261200800001E. Use of sector 22 (Southeast Region Collaborative Access team) at the Advanced Photon Source was supported by the US Department of Energy, Basic Energy Sciences, Office of Science, under contract number W-31-109-Eng-38. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government.

Author Contributions M.C., J.H., L.L., G.O., J.R.M. and P.D.K. designed the study, analysed the data, and prepared this manuscript. J.H. and L.L. performed B-cell sorting, antibody cloning, epitope mapping assay, MPER-specific neutralizing sera screening and assessed the impact of sequence variation on 10E8 neutralization. M.K.L. and J.R.M. tested the breadth and potency of 10E8. B.C., S.K.S. and R.W. performed the infected cell surface staining and antibody-virion washout assays. S.M.A. and B.F.H. performed the autoreactivity assays. G.O., Y.Y. and P.D.K. performed 10E8 structural analysis, with T.W. and B.Z. assisting with paratope alanine scanning. R.T.B. screened the B-cell culture supernatants for neutralization activity. H.I. sequenced the patient N152 virus. S.A.M. led the clinical care of the patients. M.C., L.L., N.A.D.-R. and N.S.L. optimized B-cell culture protocol.

Author Information The nucleotide sequence of 10E8 heavy and light chains have been submitted to GenBank under accession numbers JX645769 and JX645770. Coordinates and structure factors for 10E8 Fab in complex with the gp41 MPER have been deposited with the Protein Data Bank under accession code 4G6F. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.C. (mconnors@nih.gov).

METHODS

Study patients. We selected the plasma and peripheral blood mononuclear cells (PBMCs) from the HIV-1-infected patients enrolled in the National Institute of Health under a clinical protocol approved by the Investigational Review Board in the National Institute of Allergy and Infectious Diseases (NIAID-IRB). All participants signed informed consent approved by the NIAID-IRB. The criteria for enrolment were as follows: having a detectable viral load, a stable CD4 T-cell count above $400 \text{ cells } \mu\text{l}^{-1}$, being diagnosed with HIV infection for at least 4 years, and off antiretroviral treatment for at least 5 years. On the basis of the locations of current and former residences, all patients were presumed to be infected with clade B virus. Donor N152 was selected for B-cell sorting and antibody generation because his serum neutralizing activity is among the most potent and broad in our cohort. He is a slow progressor based on criteria described previously⁴¹. At the time of leukapheresis, he had been infected with HIV-1 for 20 years, with CD4 T-cell counts of $325 \text{ cells } \mu\text{l}^{-1}$, plasma HIV-1 RNA values of $3,811 \text{ copies ml}^{-1}$ and was not on antiretroviral treatment.

Viruses and plasmids. HIV-1 JR2 MPER alanine mutant pseudovirus plasmids were obtained from M. Zwick (The Scripps Research Institute). HIV-2/HIV-1 chimaeras were provided by G. Shaw and L. Morris.

Memory B-cell staining, sorting and antibody cloning. Staining and single-cell sorting of memory B cells were performed as follows. PBMCs from HIV-1-infected donor N152 were stained with antibody cocktail consisting of anti-CD19-PE-Cy7 (BD Bioscience), IgA-APC (Jackson ImmunoResearch Laboratories Inc.), IgD-FITC (BD Pharmingen) and IgM-PE (Jackson ImmunoResearch Laboratories Inc.) at 4°C in dark for 30 min. The cells were then washed with 10 ml PBS-BSA buffer and re-suspended in $500 \mu\text{l}$ PBS-BSA. A total of 66,000 $\text{CD19}^+\text{IgA}^-\text{IgD}^+\text{IgM}^+$ memory B cells were sorted using a FACSAria III cell sorter (Becton Dickinson) and re-suspended in IMDM medium with 10% FBS containing 100 U ml^{-1} IL-2, 50 ng ml^{-1} IL-21 and $1 \times 10^5 \text{ ml}^{-1}$ irradiated 3T3-MSCD40L feeder cells⁴². B cells were seeded into 384-well microtitre plates at a density of 4 cells per well in a final volume of $50 \mu\text{l}$. After 13 days of incubation, $40 \mu\text{l}$ of culture supernatants from each well were collected and screened for neutralization activity using a high-throughput micro-neutralization assay against HIV-1_{MN.3} and HIV-1_{Bal.26}. B cells in each well were lysed with $20 \mu\text{l}$ lysis buffer containing $0.25 \mu\text{l}$ of RNase inhibitor (New England Biolabs Inc.), $0.3 \mu\text{l}$ of 1 M Tris pH 8 (Quality Biological Inc.) and $19.45 \mu\text{l}$ DEPC-treated H_2O . The plates with B cells were stored at -80°C .

The variable region of the heavy chain and the light chain of the immunoglobulin gene were amplified by RT-PCR from the wells that scored positive in both the HIV-1_{MN.3} and HIV-1_{Bal.26} neutralization assay. The cDNA product was used as template in the PCR reaction. To amplify the highly somatically mutated immunoglobulin gene, two sets of primers as described previously²¹ were used in two independent PCRs. One set of primers consisted of the forward primers and the reverse primers specific for the leader region and constant region of IgH, IgK or IgL, respectively. The other set of primers consisted of the forward primer mixes specific for FWRI and respective reverse primers specific for the IgH, IgK and IgL J genes. All PCRs were performed in 96-well PCR plates in a total volume of $50 \mu\text{l}$ containing 20 nM each primer or primer mix, 10 nM each dNTP (Invitrogen), $10 \mu\text{l}$ $5\times$ Q-solution (Qiagen) and 1.2 U HotStar Taq DNA polymerase (Qiagen). From the positive PCR reactions, pools of the V_H or V_L-region DNA were ligated to a pCR2.1-Topo-TA vector (Invitrogen) for sequencing before cloning into the corresponding IgY1, IgK and IgL expression vector. $10 \mu\text{g}$ of heavy and light chain plasmids, cloned from the same well and combined in all possible heavy and light chain pairs, were mixed with $40 \mu\text{l}$ FuGENE 6 (Roche) in $1,500 \mu\text{l}$ DMEM (Gibco) and co-transfected into 293T cells. The full-length IgG was purified using a recombinant protein-A column (GE Healthcare).

Neutralization assays. Neutralization of the monoclonal antibodies was measured using single-round HIV-1 Env-pseudovirus infection of TZM-bl cells⁴³. HIV-1 Env pseudoviruses were generated by co-transfection of 293T cells with pSG3 delta Env backbone and a second plasmid that expressed HIV-1 Env. At 72 h after transfection, supernatants containing pseudovirus were harvested and frozen at -80°C until further use. In the neutralization assay, $10 \mu\text{l}$ of fivefold serially diluted patient serum or monoclonal antibody was incubated with $40 \mu\text{l}$ of pseudovirus in a 96-well plate at 37°C for 30 min before addition of TZM-bl cells. After 2 days of incubation, cells were lysed and the viral infectivity was quantified by measuring luciferase activity with a Victor Light luminometer (Perkin Elmer). The 50% inhibitory concentration (IC_{50}) was calculated as the antibody concentration that reduced infection by 50%. Antibody epitopes were mapped using HIV-1 JR2 MPER alanine mutant pseudoviruses in a TZM-bl assay.

HIV-2/HIV-1 chimaera neutralization. HIV-2/HIV-1 C1 chimaera (HIV-2 virus 7312A with HIV-1 gp41 MPER)²⁵ was used in the competition assay. A fixed concentration of MPER peptide was incubated with serially diluted 2F5,

4E10, Z13e1 or 10E8 antibody at 37°C for 30 min before incubation with HIV-2/HIV-1 C1 chimaera. Wild-type HIV-2 virus 7312A was used as a control. Antibody epitope mapping was completed by adding $10 \mu\text{l}$ 10E8 monoclonal antibody to $5 \mu\text{l}$ serial dilutions of 4E10 peptide or its alanine mutants at 37°C for 30 min before the addition of HIV-2/HIV-1 C1 chimaera. The degree to which peptides blocked antibody-mediated neutralization was calculated as the fold change in the IC_{50} value of the antibody in the presence of 4E10 alanine mutants compared to the wild-type peptide. The precise binding region within the MPER targeted by patient serum or antibodies was determined using the HIV-2/HIV-1 chimaeras containing different portions of HIV-1 MPER, such as C1 (HIV-2 Env with HIV-1 MPER), C1C (HIV-2 Env with clade C MPER), C3 (HIV-2 Env with 2F5 epitope), C4 (HIV-2 Env with 4E10 epitope), C6 (HIV-2 Env with short 4E10 epitope NWFDIT), C7 (HIV-2 Env with short 2F5 epitope ALDKWA) and C8 (HIV-2 Env with both Z13 and 4E10 epitope). Fivefold diluted patient serum or monoclonal antibody was incubated with chimaera in a 96-well plate at 37°C for 30 min before addition of TZM-bl cells. The specificities within patient sera were confirmed by blocking neutralization of the C1 chimaera with $25 \mu\text{g ml}^{-1}$ of 2F5, 4E10, MPER, Bal.V3, control peptide, or $50 \mu\text{g ml}^{-1}$ of Z13 peptide.

ELISA assays. Each antigen at $2 \mu\text{g ml}^{-1}$ was coated on 96-well plates overnight at 4°C . Plates were blocked with BLOTTO buffer (PBS, 1% FBS, 5% non-fat milk) for 1 h at room temperature, followed by incubation with antibody serially diluted in disruption buffer (PBS, 5% FBS, 2% BSA, 1% Tween-20) for 1 h at room temperature. $1:10,000$ dilution of horseradish peroxidase (HRP)-conjugated goat anti-human IgG antibody was added for 1 h at room temperature. Plates were washed between each step with 0.2% Tween 20 in PBS. Plates were developed using 3,3',5,5'-tetramethylbenzidine (TMB) (Sigma) and read at 450 nm.

Autoreactivity assays. Binding of 10E8 to phospholipid was measured by SPR conducted on a BIAcore 3000 instrument and data analyses were performed using the BIAevaluation 4.1 software (BIAcore) as described previously³⁰. Phospholipid-containing liposomes were captured on a BIAcore L1 sensor chip, which uses an alkyl linker for anchoring lipids. Before capturing lipids, the surface of the L1 chip was cleaned with a 60-s injection of 40 mM octyl- β -D-glucopyranoside, at $100 \mu\text{l min}^{-1}$, and the chip and fluidics were washed with excess buffer to remove any traces of detergent. Monoclonal antibodies were then injected at $100 \mu\text{g ml}^{-1}$ at a flow rate of $30 \mu\text{l min}^{-1}$. After each antibody injection, the surface was again cleaned with octyl- β -D-glucopyranoside, and 5-s injections of each 5 mM HCl, then 5 mM NaOH, to clean any adherent protein from the chip.

Reactivity to HIV-1 negative human epithelial (HEp-2) cells was determined by indirect immunofluorescence on slides using Evans Blue as a counterstain and FITC-conjugated goat anti-human IgG (Zeus Scientific)²⁹. Slides were photographed on a Nikon Optiphot fluorescence microscope. Regarding Fig. 3b, koda-chrome slides were taken of each monoclonal antibody binding to HEp-2 cells at a 10-s exposure, and the slides scanned into digital format. The Luminex AteNA Multi-Lyte ANA test (Wampole Laboratories) was used to test for monoclonal antibody reactivity to SSA/Ro, SS-B/La, Sm, ribonucleoprotein (RNP), Jo-1, double-stranded DNA, centromere B, and histone and was performed as per the manufacturer's specifications and as previously described²⁹. Monoclonal antibody concentrations assayed were 50, 25, 12.5 and $6.25 \mu\text{g ml}^{-1}$. $10 \mu\text{l}$ of each concentration were incubated with the luminex fluorescent beads and the test performed per the manufacturer's specifications.

Fluorescence-activated cell sorting (FACS) staining of cell-surface HIV-1 Env. FACS staining was performed as previously described^{31,44}. Forty-eight hours after transfection, cells were collected and washed in FACS buffer (PBS, 5% HIFBS, 0.02% azide) and stained with monoclonal antibodies. The transfected cells were suspended in FACS buffer and were incubated with the antibodies for 1 h at room temperature. The monoclonal antibody-cell mixture was washed extensively in FACS buffer and phycoerythrin (PE)-conjugated goat anti-human secondary antibody (Sigma) was added for 1 h at a 1:200 dilution, followed by extensive washing to remove unbound secondary antibody. The antibody-PE-stained cells were acquired on a BD LSRII instrument and analysed by FlowJo.

Antibody-virus washout experiments. From a starting concentration of 2 mg ml^{-1} , $12.5 \mu\text{l}$ of fivefold serially diluted antibodies in PBS were added to $487.5 \mu\text{l}$ of DMEM containing 10% heat-inactivated FBS and $15 \mu\text{l}$ of pseudovirus such that the final concentrations of antibodies were $50 \mu\text{g ml}^{-1}$ to $0.08 \mu\text{g ml}^{-1}$ in a total volume of $500 \mu\text{l}$. In the 'no inhibitor' control, the same volume of PBS was added instead of antibody. The reaction mixture was incubated for 30 min at 37°C . The $250 \mu\text{l}$ reaction mixture was diluted to 10 ml with complete DMEM, centrifuged at 25,000 r.p.m. in a SW41 rotor, for 2 h at 4°C . The virus pellet was then washed two additional times with 10 ml of PBS. During the washing steps, the virus-antibody complex was centrifuged at 40,000 r.p.m. for 20 min at 4°C . After the final wash, $250 \mu\text{l}$ of DMEM was added to the washed virus pellet and it was re-suspended by gentle shaking at 4°C for 30 min. A total of $100 \mu\text{l}$ of the suspended virus was used to infect $100 \mu\text{l}$ of TZM-bl cells ($0.2 \text{ million per ml}$), in

duplicate. From the remaining 250 µl of reaction mixture, an equal volume of the antibody virus mixture was used as a 'no washout' control. Plates were incubated at 37 °C in a CO₂ incubator for 2 days. After 2 days, the luciferase assay was done as described previously⁴⁵. The data were then plotted to determine the neutralization mediated by the antibodies in 'wash' or 'no wash' conditions.

Structure determination and analysis. The antigen binding fragment of 10E8 (Fab) was prepared using LysC digestion, as previously described⁴⁶. The IgG was first reduced with 100 mM dithiothreitol (DTT) for 1 h at 37 °C, followed by 1 h of dialysis in HEPES, pH 7.6, to reduce the DTT concentration to 1 mM. Antibodies were then dialysed against 2 mM iodoacetamide for 48 h at 4 °C, and subjected to a final dialysis against HEPES, pH 7.6, for 2 h. After reduction and alkylation, antibodies were cleaved with Lys-C (Roche). Fab was purified by ion exchange (Mono S) and size-exclusion chromatography (S200). Purified 10E8 Fab was incubated with tenfold excess peptide RRR-NEQELLELDKWASLWNWFDITNWLWYIR-RRR (American Peptide) and the complex then set up for 20 °C vapour diffusion sitting-drop crystallizations on a Honeybee 963 robot. A total of 576 initial conditions adapted from the commercially available Hampton (Hampton Research), Precipitant Synergy (Emerald Biosystems) and Wizard (Emerald Biosystems) crystallization screens were set up and imaged using the Rockimager (Formulatrix), followed by hand optimization of crystal hits. Crystals grown in 40% PEG 400, 0.1 M NaCitrate, 0.1 M Tris pH 7.5 diffracted to 2.1 Å resolution in a cryoprotectant composed of mother liquor supplemented with 15% 2R-3R-butanediol and excess peptide. After mounting the crystals on a loop, they were flash cooled and data collected at 1.00 Å wavelength at SER CAT ID-22 or BM-22 beamlines (APS) and processed using HKL-2000⁴⁷. Structures were solved through molecular replacement with Phaser^{48,49}, using a previously obtained free structure of 10E8 as a search model. Refinement of the structure was undertaken with Phenix⁵⁰, with iterative model building using Coot⁵¹. The structure was validated with MolProbity⁵², yielding 97% and 99.8% of residues falling within most favoured and allowed Ramachandran regions, respectively. The structure was analysed with APBS⁵³ for electrostatics, Ligplot⁵⁴ for direct contacts, PISA⁵⁵ for buried surface areas, and lsqkab (ccp4 Package⁵⁶) for r.m.s.d. alignments. Helical wheels were generated using the program Pepwheel (<http://150.185.138.86/cgi-bin/emboss/pepwheel>). All graphics were prepared with Pymol (PyMOL Molecular Graphics System).

Assessment of binding affinities of 10E8 and 10E8 variants to the gp41 MPER. Surface-Plasmon Resonance (SPR) (Biacore T200, GE Healthcare) was used to assess binding affinity of wild-type 10E8 to a gp41 MPER peptide. A biotinylated peptide composed of residues 656–683 of the gp41 MPER (RRR-NEQELLELDKWASLWNWFDITNWLWYIR-RRK-biotin; American Peptide) was coupled to a biacore SA chip to a surface density of 20–50 response units (RU). The 10E8 fragment of antigen binding (Fab) was then flowed over as analyte at concentrations ranging from 0.25 nM to 125 nM, at twofold serial dilutions, with association and dissociation phases of up to 5 min, at a flow rate of 30 µl min⁻¹. The binding of the 2F5 and 4E10 Fab controls to the same peptide was examined under identical conditions.

Binding affinities of the 10E8 paratope alanine mutants to the MPER were also assessed with SPR, but using an antibody capture method. A Biacore CM5 chip was amine-coupled with anti-human Fc antibody to high surface densities of ~10,000 RU. The 10E8 paratope variant IgGs were then captured to between 1,500–2,500 RU and a peptide composed of residues 656–683 of the gp41 MPER (RRR-NEQELLELDKWASLWNWFDITNWLWYIR-RRR) flowed over as analyte at twofold serial dilutions starting at 500 nM (with the exception of HCD30A, W100bA, S100cA, P100fA, which started at 250 nM). Association and dissociation phases spanned 3 min and 5 min, respectively, at a flow rate of 30 µl min⁻¹. Binding sensograms were fit with 1:1 Langmuir models using Biacore BioEvaluation Software (GE Healthcare). In all cases, Biacore HBSEP+ buffer was used (10 mM HEPES, pH 7.4, 150 mM NaCl, 3 mM EDTA, 0.1% P-20).

PCR amplification and sequencing. Extraction of viral RNA from plasma and cDNA synthesis were performed as previously described⁵⁷. Single molecules of a 588-bp fragment, encompassing the MPER region of the HIV-1 envelope gene, obtained through limiting dilution, were PCR-amplified with the Expand High Fidelity PCR System (Roche Applied Science) using the following primer sets: +7789 (sense) 5'-TCTTAGGAGCAGCAGGAAGCACTATGGG-3' and -8524 (antisense) 5'-GTAAGTCTCTCAAGCGGTGGTAGC-3' in a first round reaction; +7850 (sense) 5'-ACAATTATTGTCTGGTATAGTGCAACAGCA-3' and -8413 (antisense) 5'-CCACCTTCTCTTCGATTCCTTCGG-3' in a second round reaction. Each round of PCR consisted of 25 cycles, with the initial denaturation at 94 °C for 2 min, followed 25 cycles of denaturation at 94 °C for 15 s, annealing at 50 °C for 30 s, and extension at 72 °C for 1 min, with the final extension at 72 °C for 7 min. The PCR products were purified with the QIA quick PCR purification kit (Qiagen), and then cloned into pCR2.1-TOPO vector (TOPO TA Cloning kit, Invitrogen) for sequence analysis of individual molecular clones. The DNAs from 18 independent clones were sequenced with the ABI BigDye Terminator v3.1 Ready Reaction Cycle Sequencing kit (Applied Biosystems) and analysed with the ABI PRISM 3130xl Genetic Analyzer (Applied Biosystems).

Statistical analysis. The relationship between the potency of N152 patient serum and 10E8, and the relationship between 10E8 variant binding and neutralization were evaluated by the Spearman rank method.

41. Migueles, S. A. *et al.* Lytic granule loading of CD8⁺ T cells is required for HIV-infected cell elimination associated with immune control. *Immunity* **29**, 1009–1021 (2008).
42. Kershaw, M. H. *et al.* Immunization against endogenous retroviral tumor-associated antigens. *Cancer Res.* **61**, 7920–7924 (2001).
43. Li, M. *et al.* Human immunodeficiency virus type 1 env clones from acute and early subtype B infections for standardized assessments of vaccine-elicited neutralizing antibodies. *J. Virol.* **79**, 10108–10125 (2005).
44. Koch, M. *et al.* Structure-based, targeted deglycosylation of HIV-1 gp120 and effects on neutralization sensitivity and antibody recognition. *Virology* **313**, 387–400 (2003).
45. Mascola, J. R. *et al.* Human immunodeficiency virus type 1 neutralization measured by flow cytometric quantitation of single-round infection of primary human T cells. *J. Virol.* **76**, 4810–4821 (2002).
46. Ofek, G. *et al.* Elicitation of structure-specific antibodies by epitope scaffolds. *Proc. Natl Acad. Sci. USA* **107**, 17880–17887 (2010).
47. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Macromol. Crystallogr. A* **276**, 307–326 (1997).
48. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
49. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D* **67**, 235–242 (2011).
50. Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58**, 1948–1954 (2002).
51. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
52. Davis, I. W. *et al.* MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375–W383 (2007).
53. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA* **98**, 10037–10041 (2001).
54. McDonald, I. K. & Thornton, J. M. Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **238**, 777–793 (1994).
55. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
56. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D* **67**, 235–242 (2011).
57. Imamichi, H. *et al.* Human immunodeficiency virus type 1 quasi species that rebound after discontinuation of highly active antiretroviral therapy are similar to the viral quasi species present before initiation of therapy. *J. Infect. Dis.* **183**, 36–50 (2001).

The Mu transpososome structure sheds light on DDE recombinase evolution

Sherwin P. Montaña¹, Ying Z. Pigli¹ & Phoebe A. Rice¹

Studies of bacteriophage Mu transposition paved the way for understanding retroviral integration and V(D)J recombination as well as many other DNA transposition reactions. Here we report the structure of the Mu transpososome—Mu transposase (MuA) in complex with bacteriophage DNA ends and target DNA—determined from data that extend anisotropically to 5.2 Å, 5.2 Å and 3.7 Å resolution, in conjunction with previously determined structures of individual domains. The highly intertwined structure illustrates why chemical activity depends on formation of the synaptic complex, and reveals that individual domains have different roles when bound to different sites. The structure also provides explanations for the increased stability of the final product complex and for its preferential recognition by the ATP-dependent unfoldase ClpX. Although MuA and many other recombinases share a structurally conserved ‘DDE’ catalytic domain, comparisons among the limited set of available complex structures indicate that some conserved features, such as catalysis *in trans* and target DNA bending, arose through convergent evolution because they are important for function.

Mobile DNA elements are important in many aspects of biology, such as disease, evolution and the spread of antibiotic resistance, and the recombinases they encode, including MuA, are useful genetic tools^{1,2}. The DNA transposition system of bacteriophage Mu was the first to be developed *in vitro*³. MuA, many other DNA transposases, and retroviral integrases share a conserved RNaseH-like or DDE catalytic domain, named for the three Mg²⁺-binding carboxylates in their active sites⁴. Structural studies have lagged behind biochemical ones: only three family members have been co-crystallized in active, DNA-bound complexes, and only one with target DNA^{5–7}. Despite mechanistic similarities, only the catalytic domain is conserved among all of these, and their overall architectures are completely different⁴. More examples are needed to understand the diverse ways in which these enzymes harness a common catalytic domain to accomplish transposition. The richness of the known biochemistry for bacteriophage Mu, from assembly of the initial complex to targeted disassembly of the product complex, makes it a particularly informative example for structural studies.

The first steps of Mu transposition (Fig. 1) are common to many other DNA transposition systems as well as retroviral integration: (1) pairing of the mobile element's ends by the recombinase to form a ‘transpososome’ or ‘intasome’; (2) hydrolytic nicking at the bacteriophage–host junction; and (3) attack of the newly freed 3' hydroxyls on a target DNA (strand transfer), creating a new connectivity. Bacteriophage Mu uses this mechanism to form a lysogen, and to replicate when it becomes lytic. During the lytic phase, host enzymes are recruited to convert the branched product into replication forks, resulting in duplication of the entire bacteriophage genome. However, during initial lysogen formation, the ‘flanking host’ DNA (grey in Fig. 1) consists only of extra sequences appended during packaging of the bacteriophage Mu DNA into bacteriophage capsids. In this case, a poorly understood signal causes the transposase to cleave both strands at each genome end, leading to a simple insertion without replication^{8,9}.

MuA is chemically active only when incorporated into transpososomes that pair the two ends of the phage genome, and thus assembly of this complex is a regulatory step. Mu transpososomes become

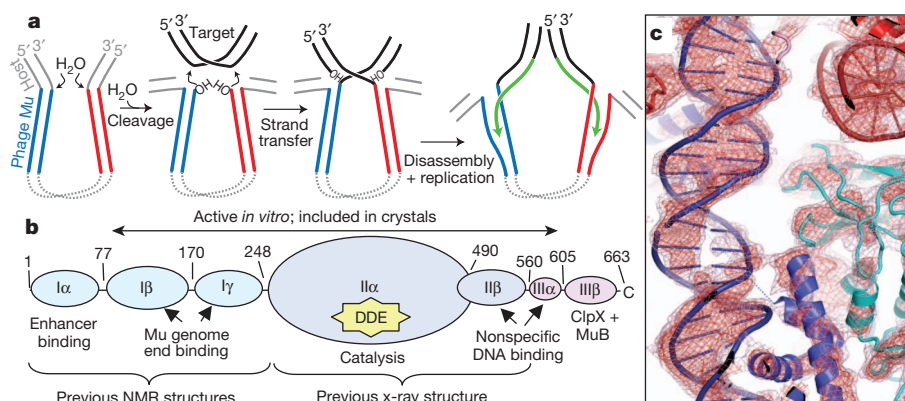


Figure 1 | Transposition pathway and structure determination. **a**, Cartoon of transposition. The transposase (MuA) pairs the bacteriophage genome ends (blue and red). At each end, the same active site catalyses the attack of H₂O at the phage–host junction and then the direct attack of the phage 3'-OH on target DNA ('strand transfer'). Target binding is nonspecific, and there is a 5-bp stagger between the sites of attack. Host and target DNAs may be entire circular replicons. After the ATP-dependent unfoldase ClpX disassembles the final strand transfer complex, the 3' hydroxyls are used as replication primers, resulting in duplication of the bacteriophage genome. Our crystals contain the strand transfer product (third panel). **b**, Domain structure of MuA. **c**, Experimental electron density map after phase improvement with Parrot superimposed on the model (contours are 1.2 and 2σ).

¹Department of Biochemistry and Molecular Biology, The University of Chicago, Chicago, Illinois 60637, USA.

increasingly stable as the reaction progresses^{10,11}. After strand transfer the complex is so stable that the enzyme MuA does not actually turn over. The strand transfer reaction can only be reversed if the complex is disrupted, for instance by heating to 75 °C (refs 12, 13). This may be a thermodynamic necessity for a reaction in which the first step (hydrolysis) is committed, yet the second step (strand transfer) is chemically isoenergetic, with no net change in the number of phosphodiester bonds. *In vivo*, Mu transpososomes must be disassembled by the ATP-dependent unfoldase ClpX before DNA replication can proceed^{14–17}. From the transpososome structure it is possible to derive explanations for the increased stability of the final complex and its preferential recognition by ClpX.

Crystallization of the strand transfer complex was facilitated by two observations. First, despite a lack of target sequence specificity, MuA attacks mismatch-containing target DNA with single-nucleotide precision (Supplementary Fig. 1)¹. Second, the natural transpososome assembly pathway (described below) can be simplified under permissive conditions *in vitro*. The resulting active complexes contain four copies of the MuA protein and two copies of a ~50-base-pair (bp) DNA derived from the bacteriophage genome's right end, each carrying two MuA binding sites (termed R1 and R2)^{18,19}. Modelling based on the structure of these complexes suggests that transpososomes formed on full left and right ends are quite similar.

Overall architecture

Viewed in isolation, the five domains of each subunit resemble beads on a string (Supplementary Fig. 2). However, when all four subunits are assembled on the DNA, they intertwine to form a network of protein–protein and protein–DNA interactions (Fig. 2 and Supplementary Video 2). The overall transpososome resembles a pair of scissors, with the bacteriophage end DNAs forming the handles and the sharply bent target DNA the blades. A 34 Å resolution electron microscopy (EM) reconstruction of Mu transpososomes in the absence of target DNA found a similar V-shape, although the arms were shorter and the accompanying electron spectroscopic imaging predicted a more contorted path for the DNA²⁰. Within the transpososome, most of the individual protein domains perform different roles in the R1- versus R2-bound subunits. Where a system encoded by a larger genome might, during the course of evolution, use two separate polypeptides, bacteriophage Mu re-uses the same sequence to perform different functions within the complex.

Catalysis and Mu DNA end-binding *in trans*

The catalytic sites lie within domain II α of MuA. In agreement with biochemical studies, only the R1-bound subunits' active sites engage with DNA, and they do so *in trans*: for example, the dark-blue subunit

binds the blue Mu DNA via domains I β and γ , whereas its active site domain docks at the red Mu DNA–target junction (the interdomain linkers are too short for any other connectivity) (Fig. 3). First characterized for Mu transposition^{21,22}, such *trans* catalysis is a recurring theme in DNA transposition, and helps to render chemical activity dependent on full complex assembly⁴. Domains II α and II β of the R2 subunits bridge the two Mu end DNAs and have a primarily structural role. II α of each R2 subunit interacts with the DNA-binding domain (DBD) I β of the subunit bound to the R1 site of the same DNA segment, whereas II β of each R2 subunit binds to the opposite DNA segment. Biochemical studies had predicted domain II β to interact with the target DNA, which does occur in the R1 subunits²³.

Domains I β and I γ of MuA recognize the specific binding sites on the bacteriophage DNA ends and their positions agree with footprinting and mutagenesis data^{24,25}. The closest structural match to MuA's tandem DNA binding domains is the centromere binding protein CENP-B, which probably evolved from an ancestral transposase²⁶. The eukaryotic mariner family Mos1 and Tc3 transposase structures also include tandem DBDs^{7,27}. In both, contacts between the DBDs equivalent to domain I β of MuA mediate synapsis of the two transposon ends. In the Mu transpososome, only the R1-bound DBDs mediate protein–protein contacts: I β as described above, and I γ to both II α of the R1 subunit at the other end of the Mu bacteriophage and III α of the R2 subunit at the same end of the Mu bacteriophage. The importance of these interactions is underscored by the high sequence conservation within the interaction surfaces (Supplementary Fig. 3).

Although the resolution precludes detailed analysis of DNA bending, the path of the backbone is clear. The R2 site is bent by ~28°, largely through compression of the major groove around domain I β , which agrees with DNase I hypersensitive sites and the CENP-B–DNA structure^{10,25,28}. The R1 site is less bent (~17°). Stronger bending there would cause a steric clash between domain I γ of the R1-bound subunit and the β -barrel (II β) of the adjacent R2 subunit. We propose that the R1 site straightens somewhat upon transpososome formation. The formation of favourable protein–DNA and protein–protein contacts within the transpososome could offset the cost of weakening contacts between DNA and domain I β .

Larger DNA conformational changes may occur on transpososome formation, as indicated by solution experiments that found ~90° bends in monomeric MuA–DNA complexes²⁹. In monomeric MuA, domain II α could interact *in cis* with its own I β domain and additional DNA bending might be induced by electrostatic interactions with II β and III α . A different but stable monomer conformation would raise the energy barrier to spontaneous tetramer formation, making it more amenable to regulation. It could also prevent premature encounters between the active site and the DNA.

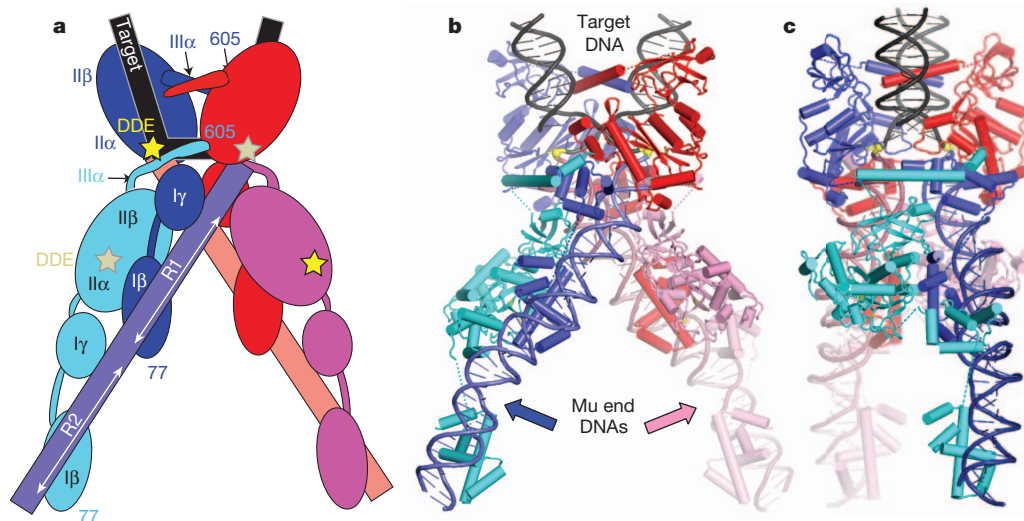


Figure 2 | Transpososome structure. The complex sits on a crystallographic two-fold symmetry axis (vertical) that relates the blue and red halves. The pale- and dark-coloured subunits adopt different conformations within the homotetramer. DNA colours match Fig. 1. **a**, Cartoon. Catalytic sites are marked as yellow and tan stars (facing the viewer or the background, respectively) and domains of the blue subunits are labelled. **b**, Ribbon drawing, with the scissile phosphate groups shown as yellow spheres. **c**, Same drawing as in **b**, rotated ~90° about a vertical axis.

treated with a high salt wash, an active complex remains that contains only four subunits of MuA, contacting only the R1, R2 and L1 DNA sites^{10,19}. Thus, the other binding sites are important for assembly but not for the final activity.

Modelling showed that the functional part of the tetrameric assembly, where Mu ends join to target DNA, can be identical in the crystallized (R1R2)₂ complex and in the full left plus right complex (Fig. 4). Subunits R1, R2, L1 and domains II and III of the L2 subunit were modelled directly from the crystal structure. The HU-induced bend allows a single protomer of MuA to bind the L2 site via domain Iβ while its domains II and III form part of the core complex. Domain Iγ acts as a linker, which explains why L2 is the only site where Iγ doesn't bind DNA²⁵.

Additional interactions involving the L3- and R3-bound subunits may temporarily hold the components together while the intertwining of protein and DNA needed to form a transpososome at the Mu–host junction occurs. We modelled interactions between the L3 subunit's domain IIα and R2's Iβ based on those seen in the crystal between the R2- and R1-bound subunits. The R3 subunit's role is unclear, but if the HU-induced bend were relaxed, similar cross-end interactions might occur between the L3 and R3 subunits.

Topological studies predicted two right-handed superhelical crossings within the transpososome^{39,40}. One such crossing occurs in the crystal structure, near the junction with target DNA (R1 over L1 in Fig. 4). A second crossing (L3 over R2) results from the severe bend induced by HU in the model, in conjunction with a smaller bend in the L2 site. Restraint of two supercoils within such short segments of DNA is consistent with observations that supercoiling stimulates transpososome assembly⁴¹.

During assembly on intact bacteriophage DNA, domain Iα transiently binds an internal enhancer, which contains two clusters of MuA binding sites termed O1 and O2. There are too many degrees of freedom to add the enhancer to our model. However, our model does agree with data showing that O1-bound proteins interact with both L3 and R1 and that proteins bound to the somewhat longer O2 may contact both R3 and L1^{42,43}. The multiple protein–protein interaction surfaces of MuA may help to stabilize an initial pairing of the bacteriophage ends, but interactions between the wrong partners could slow down the transition to a final, cleavage-ready complex. The enhancer may stimulate this transition by preventing unproductive interactions among L- and R-bound subunits as well as by aligning them for productive ones.

Such an elaborate assembly process is not limited to Mu: many other mobile elements also require seemingly 'extra' recombinase subunits⁴⁴. Although the details vary among these systems, they may all be using the same fundamental strategy of using additional subunits to temporarily stabilize pairing of the element end DNAs while a complicated, intertwined structure forms at the 3' ends. The additional complexity may also provide additional opportunities for regulation. Finally, for mobile elements that are present in high copy, it may help to ensure that the two ends paired in a single transpososome belong to the same copy of the mobile element.

Convergent and divergent evolution

Many DNA transposases and retroviral integrases share a structurally conserved DDE or RNaseH-like catalytic domain, indicative of divergence from a common ancestor. However, this is the only domain conserved among these diverse recombinases. Comparison of the four reported structures of DDE recombinases in complex with substrate DNAs shows that other recurring features may reflect convergent evolution for functional reasons (Fig. 5). All four complexes are held together by intertwined networks of protein–protein and protein–DNA contacts, although different domains mediate those contacts⁴. Mos1 and MuA do have structurally related bipartite DNA-binding domains, but even those domains form different protein–protein contacts in their respective transpososomes⁷.

Despite the diversity of these complexes, catalysis is always in *trans*: the subunit that catalyses DNA cleavage and joining on one mobile element

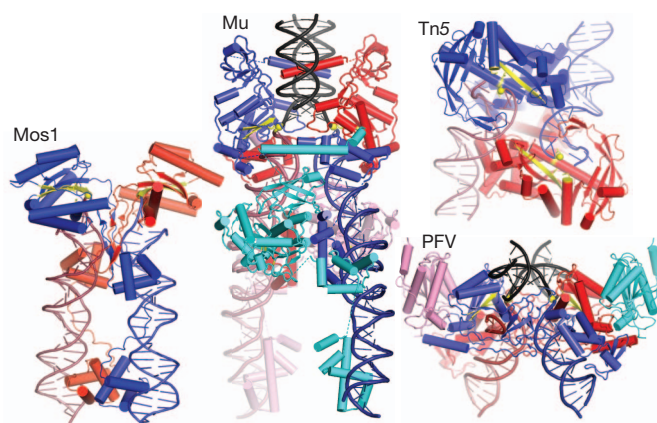


Figure 5 | Comparison of DDE recombinase–DNA complexes. The mobile element ends are red and blue, and target DNA (where included) is black. Subunits that carry out the chemical reactions are red and blue; additional subunits are pink and cyan. Active site residues, scissile phosphate groups, and the two β-strands of the conserved catalytic domain that carry the catalytic D residues are in yellow. Mos1 is a Tc1/mariner family eukaryotic DNA transposon; Tn5 is a bacterial DNA transposon; and PFV is a mammalian retrovirus^{5–7}. Mos1 and Tn5 require only a dimer for activity, whereas Mu transposase and PFV integrase require tetramers. In the PFV structure, only the catalytic domains of the additional subunits were visible (pink and cyan).

end binds to specific sequences on the other end. This feature ensures that the chemical reactions at the two element ends are coordinated, because the complex requires proper pairing of the ends for assembly.

Another recurring feature is strong bending of the target DNA. Target DNA bound by the prototype foamy virus (PFV) intasome is also bent, although not quite as severely as that in the Mu transpososome⁶. The Mos1 transpososome was crystallized without target DNA, but additional end DNAs found in the crystal bind where target is expected to, and in a way that requires target bending⁷. Modelling of target DNA onto the Tn5 transpososome structure also requires bending, which agrees with biochemical data for the related Tn10 system^{5,45}. Outside of the catalytic domain, contacts to the target DNA vary widely among these structures. Why then have they all evolved to strongly bend the target DNA? As noted for the PFV structure, target bending may help render strand transfer irreversible by straining the DNA conformation such that the ends snap away from the active site after strand transfer. This may be a source of the product binding energy that drives forward the otherwise-isoenergetic strand transfer reaction. The overall conformation of the target DNA in the Mu transpososome resembles that bound by integration host factor (IHF). In that case, a nick at the kink does enhance affinity by allowing the ends to spring apart⁴⁶.

The DDE catalytic domain is thus a conserved module that has been co-opted by numerous mobile elements to perform similar chemical reactions. However, other similarities in the way that it has been harnessed to mobilize these elements seem to reflect convergent evolution to satisfy functional requirements.

METHODS SUMMARY

We determined the structure of the final strand transfer complex, which contains a tetramer of MuA, two copies of the bacteriophage end DNA, and one target DNA (Supplementary Fig. 1 and Supplementary Video 1). Crystallizations used a slightly truncated protein, MuA(77–605), which is active *in vitro* and lacks only the N-terminal enhancer-binding domain and the C-terminal domain that interacts with ClpX and MuB, a second bacteriophage-encoded protein that helps deliver an appropriate target DNA under non-permissive conditions⁴⁷. The bacteriophage end DNA mimics pre-cleaved right ends, and the 35-bp target DNA contains a central G•G mismatch. Although MuA displays little sequence specificity for target DNA, it attacks mismatch-containing DNA with single-nucleotide precision¹. This feature facilitated production of a homogeneous sample for crystallization.

Phases were determined by MIRAS using three derivatives. The crystals diffracted anisotropically to 5.2 Å, 5.2 Å and 3.7 Å along the three principal axes.

Model building was possible despite the low resolution because >90% of the protein structure had been previously determined as isolated domains^{31,48,49}. Placement of the protein domains was verified by SeMet data, and the DNA sequence register by an additional data set collected from crystals where every T on one strand had been substituted with 5-bromodeoxyuridine (BrdU) (Supplementary Fig. 4). No previous structure was available for domain III α , which comprises ~45 amino acids that are strongly predicted to form one long helix followed by a short one. Although density for these helices was visible, the sequence register is uncertain. The complex lies on a crystallographic two-fold axis such that the asymmetric unit contains half a transpososome. After highly restrained refinement, *R* and *R*_{free} values were 39.3% and 43.7%, well within the range expected for a low-resolution structure.

Full Methods and any associated references are available in the online version of the paper.

Received 9 June; accepted 19 September 2012.

Published online 7 November 2012.

- Yanagihara, K. & Mizuuchi, K. Mismatch-targeted transposition of Mu: a new strategy to map genetic polymorphism. *Proc. Natl Acad. Sci. USA* **99**, 11317–11321 (2002).
- Haapa, S., Taira, S., Heikkinen, E. & Savilahti, H. An efficient and accurate integration of mini-Mu transposons *in vitro*: a general methodology for functional genetic analysis and molecular biology applications. *Nucleic Acids Res.* **27**, 2777–2784 (1999).
- Mizuuchi, K. *In vitro* transposition of bacteriophage Mu: a biochemical approach to a novel replication reaction. *Cell* **35**, 785–794 (1983).
- Montaño, S. P. & Rice, P. A. Moving DNA around: DNA transposition and retroviral integration. *Curr. Opin. Struct. Biol.* **21**, 370–378 (2011).
- Davies, D. R., Goryshin, I. Y., Reznikoff, W. S. & Rayment, I. Three-dimensional structure of the Tn5 synaptic complex transposition intermediate. *Science* **289**, 77–85 (2000).
- Maertens, G. N., Hare, S. & Cherepanov, P. The mechanism of retroviral integration from X-ray structures of its key intermediates. *Nature* **468**, 326–329 (2010).
- Richardson, J. M., Colloms, S. D., Finnegan, D. J. & Walkinshaw, M. D. Molecular architecture of the Mos1 paired-end complex: the structural basis of DNA transposition in a eukaryote. *Cell* **138**, 1096–1108 (2009).
- Choi, W. & Harshey, R. M. DNA repair by the cryptic endonuclease activity of Mu transposase. *Proc. Natl Acad. Sci. USA* **107**, 10014–10019 (2010).
- Chaconas, G., Kennedy, D. L. & Evans, D. Predominant integration end products of infecting bacteriophage Mu DNA are simple insertions with no preference for integration of either Mu DNA strand. *Virology* **128**, 48–59 (1983).
- Lavoie, B. D., Chan, B. S., Allison, R. G. & Chaconas, G. Structural aspects of a higher order nucleoprotein complex: induction of an altered DNA structure at the Mu-host junction of the Mu type 1 transpososome. *EMBO J.* **10**, 3051–3059 (1991).
- Surette, M. G., Buch, S. J. & Chaconas, G. Transpososomes: stable protein-DNA complexes involved in the *in vitro* transposition of bacteriophage Mu DNA. *Cell* **49**, 253–262 (1987).
- Au, T. K., Pathania, S. & Harshey, R. M. True reversal of Mu integration. *EMBO J.* **23**, 3408–3420 (2004).
- Mizuuchi, M., Rice, P. A., Wardle, S. J., Haniford, D. B. & Mizuuchi, K. Control of transposase activity within a transpososome by the configuration of the flanking DNA segment of the transposon. *Proc. Natl Acad. Sci. USA* **104**, 14622–14627 (2007).
- Krukltis, R., Welty, D. J. & Nakai, H. ClpX protein of *Escherichia coli* activates bacteriophage Mu transposase in the strand transfer complex for initiation of Mu DNA synthesis. *EMBO J.* **15**, 935–944 (1996).
- Levchenko, I., Luo, L. & Baker, T. A. Disassembly of the Mu transposase tetramer by the ClpX chaperone. *Genes Dev.* **9**, 2399–2408 (1995).
- Mhammedi-Alaoui, A., Pato, M., Gama, M. J. & Toussaint, A. A new component of bacteriophage Mu replicative transposition machinery: the *Escherichia coli* ClpX protein. *Mol. Microbiol.* **11**, 1109–1116 (1994).
- Abdelhakim, A. H., Oakes, E. C., Sauer, R. T. & Baker, T. A. Unique contacts direct high-priority recognition of the tetrameric Mu transposase-DNA complex by the AAA+ unfoldase ClpX. *Mol. Cell* **30**, 39–50 (2008).
- Savilahti, H., Rice, P. A. & Mizuuchi, K. The phage Mu transpososome core: DNA requirements for assembly and function. *EMBO J.* **14**, 4893–4903 (1995).
- Baker, T. A. & Mizuuchi, K. DNA-promoted assembly of the active tetramer of the Mu transposase. *Genes Dev.* **6**, 2221–2232 (1992).
- Yuan, J. F., Beniac, D. R., Chaconas, G. & Ottensmeyer, F. P. 3D reconstruction of the Mu transposase and the Type 1 transpososome: a structural framework for Mu DNA transposition. *Genes Dev.* **19**, 840–852 (2005).
- Savilahti, H. & Mizuuchi, K. Mu transpositional recombination: donor DNA cleavage and strand transfer *in trans* by the Mu transposase. *Cell* **85**, 271–280 (1996).
- Aldaz, H., Schuster, E. & Baker, T. A. The interwoven architecture of the Mu transposase couples DNA synapsis to catalysis. *Cell* **85**, 257–269 (1996).
- Kremontsova, E., Giffin, M. J., Pincus, D. & Baker, T. A. Mutational analysis of the Mu transposase. Contributions of two distinct regions of domain II to recombination. *J. Biol. Chem.* **273**, 31358–31365 (1998).
- Namgong, S. Y., Sankaralingam, S. & Harshey, R. M. Altering the DNA-binding specificity of Mu transposase *in vitro*. *Nucleic Acids Res.* **26**, 3521–3527 (1998).
- Zou, A. H., Leung, P. C. & Harshey, R. M. Transposase contacts with mu DNA ends. *J. Biol. Chem.* **266**, 20476–20482 (1991).
- Tanaka, Y. *et al.* Crystal structure of the CENP-B protein-DNA complex: the DNA-binding domains of CENP-B induce kinks in the CENP-B box DNA. *EMBO J.* **20**, 6612–6618 (2001).
- Watkins, S., van Pouderoyen, G. & Sixma, T. K. Structural analysis of the bipartite DNA-binding domain of Tc3 transposase bound to transposon DNA. *Nucleic Acids Res.* **32**, 4306–4312 (2004).
- Craigie, R., Mizuuchi, M. & Mizuuchi, K. Site-specific recognition of the bacteriophage Mu ends by the Mu A protein. *Cell* **39**, 387–394 (1984).
- Kuo, C. F., Zou, A. H., Jayaram, M., Getzoff, E. & Harshey, R. DNA-protein complexes during attachment-site synapsis in Mu DNA transposition. *EMBO J.* **10**, 1585–1591 (1991).
- Mizuuchi, M., Baker, T. A. & Mizuuchi, K. DNase protection analysis of the stable synaptic complexes involved in Mu transposition. *Proc. Natl Acad. Sci. USA* **88**, 9031–9035 (1991).
- Rice, P. & Mizuuchi, K. Structure of the bacteriophage Mu transposase core: a common structural motif for DNA transposition and retroviral integration. *Cell* **82**, 209–220 (1995).
- Wu, Z. & Chaconas, G. A novel DNA binding and nuclease activity in domain III of Mu transposase: evidence for a catalytic region involved in donor cleavage. *EMBO J.* **14**, 3835–3843 (1995).
- Abdelhakim, A. H., Sauer, R. T. & Baker, T. A. The AAA+ ClpX machine unfolds a keystone subunit to remodel the Mu transpososome. *Proc. Natl Acad. Sci. USA* **107**, 2437–2442 (2010).
- Burton, B. M. & Baker, T. A. Mu transpososome architecture ensures that unfolding by ClpX or proteolysis by ClpXP remodels but does not destroy the complex. *Chem. Biol.* **10**, 463–472 (2003).
- Naigamwalla, D. Z., Coros, C. J., Wu, Z. & Chaconas, G. Mutations in domain III α of the Mu transposase: evidence suggesting an active site component which interacts with the Mu-host junction. *J. Mol. Biol.* **282**, 265–274 (1998).
- Yang, J. Y., Kim, K., Jayaram, M. & Harshey, R. M. A domain sharing motif for active site assembly within the Mu A tetramer during transposition: the enhancer may specify domain contributions. *EMBO J.* **14**, 2374–2384 (1995).
- Surette, M. G. & Chaconas, G. The Mu transpositional enhancer can function *in trans*: requirement of the enhancer for synapsis but not strand cleavage. *Cell* **68**, 1101–1108 (1992).
- Mizuuchi, M. & Mizuuchi, K. Conformational isomerization in phage Mu transpososome assembly: effects of the transpositional enhancer and of MuB. *EMBO J.* **20**, 6927–6935 (2001).
- Harshey, R. M. & Jayaram, M. The mu transpososome through a topological lens. *Crit. Rev. Biochem. Mol. Biol.* **41**, 387–405 (2006).
- Craigie, R. & Mizuuchi, K. Role of DNA topology in Mu transposition: mechanism of sensing the relative orientation of two DNA segments. *Cell* **45**, 793–800 (1986).
- Surette, M. G. & Chaconas, G. A protein factor which reduces the negative supercoiling requirement in the Mu DNA strand transfer reaction is *Escherichia coli* integration host factor. *J. Biol. Chem.* **264**, 3028–3034 (1989).
- Allison, R. G. & Chaconas, G. Role of the A protein-binding sites in the *in vitro* transposition of Mu DNA. A complex circuit of interactions involving the Mu ends and the transpositional enhancer. *J. Biol. Chem.* **267**, 19963–19970 (1992).
- Jiang, H., Yang, J. Y. & Harshey, R. M. Criss-crossed interactions between the enhancer and the att sites of phage Mu during DNA transposition. *EMBO J.* **18**, 3845–3855 (1999).
- Craig, N. L. *Mobile DNA II* (ASM Press, 2002).
- Pribil, P. A. & Haniford, D. B. Target DNA bending is an important specificity determinant in target site selection in Tn10 transposition. *J. Mol. Biol.* **330**, 247–259 (2003).
- Swinger, K. K. & Rice, P. A. Structure-based analysis of HU-DNA binding. *J. Mol. Biol.* **365**, 1005–1016 (2007).
- Levchenko, I., Yamauchi, M. & Baker, T. A. ClpX and MuB interact with overlapping regions of Mu transposase: implications for control of the transposition pathway. *Genes Dev.* **11**, 1561–1572 (1997).
- Clubb, R. T., Schumacher, S., Mizuuchi, K., Gronenborn, A. M. & Clore, G. M. Solution structure of the Iy subdomain of the Mu end DNA-binding domain of phage Mu transposase. *J. Mol. Biol.* **273**, 19–25 (1997).
- Schumacher, S. *et al.* Solution structure of the Mu end DNA-binding I β subdomain of phage Mu transposase: modular DNA recognition by two tethered domains. *EMBO J.* **16**, 7532–7541 (1997).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank K. Mizuuchi for initiating this project, K. K. Swinger and B. Vertessy for early crystallization efforts, and X. Yang and the staff of APS beamlines 14, 19 and 21 for assistance with data collection. This work was funded in part by NIH grant GM086826 (to P.A.R.).

Author Contributions S.P.M. carried out most of the crystallographic work, Y.Z.P. grew the first diffracting transpososome crystals and assisted with all other aspects of the project, and P.A.R. designed the project and assisted in computational work and interpretation of the results.

Author Information Coordinates and structure factors were deposited at the Protein Data Bank under accession 4fcy. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.A.R. (P.Rice@uchicago.edu).

METHODS

Overview. We determined the structure of the final strand transfer complex, which contains a tetramer of MuA, two copies of the bacteriophage end DNA, and one target DNA (Fig. 2 and Supplementary Video 1). Crystallizations used a slightly truncated protein, MuA(77–605), which is active *in vitro* and lacks only the N-terminal enhancer-binding domain and the C-terminal domain that interacts with ClpX and MuB, a second bacteriophage-encoded protein that helps deliver an appropriate target DNA under non-permissive conditions⁴⁷. The bacteriophage end DNA mimics pre-cleaved right ends, and the 35-bp target DNA contains a central G•G mismatch. Although MuA displays little sequence specificity for target DNA, it attacks mismatch-containing DNA with single-nucleotide precision¹. This feature facilitated production of a homogenous sample for crystallization.

Phases were determined by MIRAS using three derivatives. The crystals diffracted anisotropically to 5.2 Å, 5.2 Å and 3.7 Å along the three principal axes. Model building was possible despite the low resolution because >90% of the protein structure had been previously determined as isolated domains^{31,48,49}. Placement of the protein domains was verified by SeMet data and the DNA sequence register by an additional data set collected from crystals where every T on one strand had been substituted with BrdU (Supplementary Fig. 4). No previous structure was available for domain III α , which comprises ~45 amino acids that are strongly predicted to form one long helix followed by a short one. Although density for these helices was visible, the sequence register is uncertain. The complex lies on a crystallographic two-fold axis such that the asymmetric unit contains half a transpososome. After highly restrained refinement, *R* and *R*_{free} values were 39.3% and 43.7%, well within the range expected for a low-resolution structure.

Expression and purification of the MuA transposase. The pMK599 plasmid, a pET3c derivative that contains the *MuA* open reading frame coding for residues 77–605, was a gift from the Mizuuchi laboratory⁵⁰. This plasmid was transformed into *Escherichia coli* Rosetta pLysS strain (EMD Biosciences) for protein overexpression. After plating transformants, a starter culture was prepared by inoculating multiple colonies into LB media (with 100 μ g ml⁻¹ ampicillin) and growing at 37 °C until the absorbance at 600 nm (*A*₆₀₀) was ~0.7. Typically, 100 ml of starter culture was prepared per litre of final culture. After the addition of starter culture to fresh ampicillin-containing LB media, cells were grown to *A*₆₀₀ of ~0.8, then protein expression was induced with IPTG (added to a final concentration of 0.5 mM). Cells were collected 2 h after induction by centrifugation at ~8,000g for 10 min, and cell pellets were stored at -80 °C for later use.

Cell pellets were re-suspended in a lysis buffer (25 mM HEPES (pH 7.50), 1 mM EDTA, 1 M NaCl, 10% sucrose, 10% glycerol, 5 mM DTT, 200 μ g ml⁻¹ lysozyme, protease inhibitor cocktail from Roche Diagnostics), sonicated, and centrifuged at 40,000g for 1 h (18,000 r.p.m. in a SS-34 rotor). Ammonium sulphate was added to the supernatant to 30% saturation to precipitate the protein. The pellet was collected by centrifugation, and redissolved in buffer A (20 mM MES (pH 5.5), 0.5 mM EDTA, 5% glycerol, 0.2 M NaCl and 1 mM DTT). The protein sample was filtered before loading onto a heparin affinity column (GE Healthcare). Proteins were eluted with salt gradient from 0.2 M to 2.0 M NaCl. To improve the purity, MuA-containing fractions were rechromatographed on heparin after dialysis into buffer A. The protein was then dialysed into buffer A again and loaded onto a Mono-S column (GE Healthcare). A gradient similar to that from the heparin affinity purification was applied. Fractions containing MuA were pooled and dialysed at 4 °C into 20 mM HEPES (pH 7.5), 0.5 mM EDTA, 0.2 M ammonium sulphate, 20% glycerol, and 1 mM DTT. The protein was concentrated to approximately 10 mg ml⁻¹, and stored at -80 °C. Minimal nuclease contamination was detected when samples (0.5 mg ml⁻¹ final concentration) were incubated for 2 h at 37 °C in 10 mM HEPES pH 7.5 with supercoiled plasmid DNA, 50 mM NaCl and 10 mM MgCl₂.

SeMet-labelled MuA(77–605) was prepared similarly except that cells were grown differently⁵¹. Instead of using LB, the cells were inoculated in M9 media plus 0.4% glucose, 10 mM NaCl, 0.1 mM CaCl₂, 2 mM MgSO₄ and 100 mg ml⁻¹ ampicillin until *A*₆₀₀ reached ~0.5. An amino acid cocktail containing L-isoleucine, L-leucine, L-lysine, L-phenylalanine, L-threonine and L-valine was then added to a final concentration of 100 mg of each amino acid per litre. Seleno-DL-methionine (Sigma) was also added to a final concentration of 60 mg l⁻¹. The culture was grown for 15 more minutes before 0.5 mM IPTG was added. Cells were collected after 3 h of induction.

Preparation of Mu end and target DNA. Mu end DNA duplexes were designed to contain the R1 and R2 binding sites for MuA. Each duplex was prepared by mixing four single strands in equal molar amounts. The oligonucleotides used for the structure determination are listed below: TL, 5'-GCTTGAAGCGGCGCA CGAAAAACGCG-3'; TR, 5'-AAAGCGTTTACGATAAATGCGAAAAC-3'; BL, 5'-AACGCTTTCGCGTTTTTCGTGCGCCGCTTCA-3'; BR, 5'-CGGTT TTCGCATTTATCGTGA-3'. These strands were heated at 80 °C for 20 min, and annealed by slow cooling to room temperature. The final concentration of

the duplex DNA is 0.2 mM in TEN buffer (10 mM Tris-HCl and 0.5 mM EDTA, 100 mM NaCl, pH 8.0). The resulting DNA mimics the product of initial DNA cleavage by MuA, and has a three-nucleotide 5'-overhang on the uncleaved strand, and a two-nucleotide 5'-overhang on the other end. Each strand of the resulting duplex is nicked at a position that does not interfere with transpososome assembly.

The target DNA contains a central mismatch and was designed to be asymmetric to avoid hairpin formation during annealing. The target DNA was prepared in a manner similar to that of the Mu end DNA using the following oligonucleotides: (1) 5'-TATCGCAACAACACATCGGATAACCATAAGTAA TA-3'; (2) 5'-TATTACTTATGGTTATCGGATGTGTTGTTGCGATA-3'.

All unmodified oligonucleotides were obtained from IDT Technologies. Brominated oligonucleotides (discussed below) that were used in this study to validate the sequence of the donor DNA and the location of the target DNA were obtained from Yale University's Keck Facility. Because brominated oligonucleotides are photolabile, they were handled in the dark as much as possible.

Crystallization and data collection. Strand-transfer complexes were assembled by mixing the target DNA, Mu end DNA, and MuA protein in 1:1.4:3.7 molar ratios in a solution containing 10 mM MgCl₂, 25 mM HEPES (pH 7.5), 10 mM DTT, 0.02% Zwittergent, 14% glycerol, and 0.2 M (NH₄)₂SO₄. This was incubated for at least 1 h at room temperature to ensure completeness of the strand transfer reaction. Although DMSO is usually added to stimulate assembly of transpososomes with two right end DNAs, we found that it was not necessary at the high protein and DNA concentrations used for crystallographic work¹⁸. Crystallization trials were then performed using the hanging-drop vapour diffusion method: drops contained a 1:1 mixture of complex stock solution (~2 mg ml⁻¹) and well, and were incubated at 19 °C. Crystals appeared in 22–28% (v/v) PEG400, 0.1 M HEPES (pH 7.50), and 0.2 M MgCl₂, and grew to their full size in 2–4 weeks. Tantalum bromide derivatives were obtained by soaking the crystals of the strand-transfer complex for 1–8 days with 0.4 mM [Ta₆Br₁₂]²⁺ cluster (Jena Bioscience) in a solution that mimicked the condition of the drop. For derivatization with mercury, crystals were soaked in 32% PEG400, 0.1 M HEPES (pH 7.5), 0.2 M MgCl₂ and 0.1 mM mersalyl acid (Sigma) for 1–2 days. For derivatization with selenium, crystallization setups were done with SeMet protein. And for that with bromine, the brominated donor DNA where every thymine on the T-rich strand (oligos BL & BR) was replaced with BrdU was used. All crystals were frozen in liquid N₂ directly from the drop.

Numerous data sets were collected from several different beamlines at the Advanced Photon Source in Argonne. Many crystals were screened at BIO-CARS 14-BM. All data sets used for the final phasing and refinement were collected at SBC-CAT 19-ID beamline at 100K temperature. For the SeMet data, data sets collected from two different crystals were merged to improve completeness of data, especially at the low-resolution shells. X-ray data collected from the native and Ta-derivative crystals were integrated and scaled with HKL3000 suite and the others with HKL2000⁵². A summary of the data collection statistics is shown in Supplementary Table 1.

Structure determination and refinement. The toehold in solving the structure of the Mu transpososome was a single Tantalum bromide cluster (Supplementary Table 1). This cluster was initially found using direct methods in SHELXD⁵³ from a 5-day-soaked Ta-data set, and was consistent with the anomalous difference Patterson maps⁵⁴ generated from other Ta-data sets where crystals were soaked for 1, 3, 5, 7 and 8 days. SIRAS phases from this one cluster were generated using MLPHARE and were used in anomalous difference Fourier methods to determine the substructure for the mercury derivative⁵⁵. We used SIRAS phases calculated from the Hg derivative to confirm independently the Ta site. After several rounds of difference Fourier calculation, we were able to locate the rest of the heavy atoms. Final MIRAS phases were generated from 4 [Ta₆Br₁₂]²⁺, 3 Hg and 17 Se sites. Reasonable figures of merit were obtained before density modification: 0.41 for centric and 0.25 for acentric reflections. With 77% solvent in the crystal, further phase improvement was achieved by density modification using Parrot⁵⁶. Electron density maps generated showed clear density for the DNA as well as tubular densities that represent protein helices. The protein structure was initially modelled by docking previously determined structural domains (I β , I γ and catalytic domains) of the MuA transposase into the density using Se peaks as markers. The bromine sites, despite not being included in calculating phases due to the low resolution of that data set, were particularly useful in guiding the model building for the donor DNA. We also have a low-resolution data set from a crystal that contains a symmetric brominated target DNA: 5'-TATCGCAACAACACA TCGGATGTGTTGTTGCGATA-3'. Bromine peaks obtained from this particular crystal was useful in confirming the location of our target DNA (Supplementary Fig. 4).

The transpososome lies on a crystallographic two-fold axis such that the asymmetric unit comprises half of a transpososome: two MuA protomers, one Mu end DNA, and

half target DNA. The initial model revealed possible loose contacts between two crystallographically related copies of domain II β (a β -barrel). To improve the diffraction of our crystals, we engineered three sets of mutations (Quickchange, Stratagene) in that region of contact: a single M521W mutation and two double mutations, M521W/N525L and M521L/N525L. The latter led to an improvement of the resolution along the best diffracting axis of our 'native data set' from 4.2 to 3.7 Å.

The structure was modelled in COOT⁵⁷, and refined using PHENIX⁵⁸. Density was visible for several sections that unfortunately could not be modelled due to the resolution: for example, the linker between domains I β and I γ lies in the minor groove as seen for Mos1 and CENP-B, and the region around the third active site residue of the R1-bound subunits, which clearly changes conformation from the inactive form seen in isolated domain II structures. Multiple restraints were used during refinement due to the low resolution of the data. These include H-bond restraints on the DNA base pairs, secondary structure restraints, model restraints where models of the individual domains of the MuA transposase were obtained from the PDB, NCS restraints, and Ramachandran restraints. Nine TLS groups were used: (1) R1 β and the DNA with which it is interacting; (2) R2 β and DNA; (3) R1 γ and DNA; (4) R2 γ and DNA plus R2-domain III α ; (5) R2-domain II including the β -barrel; (6) R1-domain II without the β -barrel; (7) R1- β -barrel; (8) R1-domain III α ; and (9) target DNA including the sequences after the CA step in the donor DNA (that do not interact directly with the DNA binding domains). As categorized by PROCHECK, the percentage of residues in the following regions of the Ramachandran plot were: favoured/allowed/generous/disallowed = 93.8/5.4/0.2/0.5. Several variations on this protocol were tried. Simply removing the Ramachandran restraints made very little difference, probably because most of the model was already restrained to previously determined domain structures. The unrestrained Ramachandran plot statistics were: favoured/allowed/generous/disallowed = 90.1/9.0/0.4/0.5, and $R_{\text{work}}/R_{\text{free}}$ were 40.1/43.6% (as opposed to 39.3/43.7% with restraints). Superimposing the two structures (refined \pm Rama restraints) revealed some slight differences in residues 347–356. This is the region where Ramachandran outliers were observed. However, upon inspection of the experimental map and the difference maps, it was difficult to discern which was more correct. Hence, we are choosing to report a structure that has better geometry. We also tried refinement with DEN⁵⁹, but it only improved the R_{free} by 0.3% and greatly degraded the Ramachandran plot. Our structure may be an unusual test case for DEN because of the low resolution of our data and the high quality of our individual domain models.

During the initial rounds of refinement, progress stalled when R_{free} was \sim 49%. However, after ellipsoidal truncation and anisotropic scaling were performed on the native data set using the Diffraction Anisotropy server⁶⁰, the R_{free} considerably improved to \sim 44%. The server truncated the data set to 3.7 Å along c^* and 5.2 Å along a^* and b^* . The final refined structure has an R_{work} and R_{free} of 39.30% and 43.70%, respectively.

Modelling the full transpososome. To model the full complex, additional model B-form DNA coordinates was created using the W3DNA server⁶¹. DNA and protein coordinates were manipulated in both pymol and coot. Subunits R1,

R2, L1 and domains II and III of the L2 subunit in the model could be taken directly from the crystal structure. To model the other subunits, we docked domain I β of the R2 subunit and the DNA segment it binds onto the appropriate site in the modelled DNA. The L1 and L2 binding sites are separated by an \sim 80-bp segment where the DNA bending protein HU binds. Modelling of the HU-induced bend was based on the structure of a closely related IHF–DNA complex and on footprinting data for HU synergistically bound within this loop^{62,63}. Modelled DNA for the L end was broken and appropriate sections abutted to the ends of the DNA in the IHF–DNA structure. We justified some additional bending of the DNA on the L2 end of the HU site based on the symmetry-related DNA in the IHF structure, and the fact that IHF- and HU-induced bends are known to be flexible. Bending of the model DNA in the L2 binding site was based on bending seen crystallographically in the R2 site. In modelling the R3-bound subunit, the other domains simply followed I β as a rigid body, which gives only a rough placement of domain II. For the L3-bound subunit, we modelled an interaction between its domain II and the R2 subunit's domain I β based on the II–I β interactions seen in the crystal.

Figures were prepared using Pymol (The PyMOL Molecular Graphics System, Version 1.3).

50. Baker, T. A., Mizuuchi, M., Savilahti, H. & Mizuuchi, K. Division of labor among monomers within the Mu transposase tetramer. *Cell* **74**, 723–733 (1993).
51. Ducruix, A. & Gieg, R. in *Preparation of Selenomethionyl Protein Crystals* (eds Dublie, S. & Carter, C. W.) (Oxford Univ. Press, 1992).
52. Otwinowski, Z. & Minor, W. in *Methods in Enzymology* Vol. 276 (eds Carter, C. W. & Sweet, R. M.) 307–326 (Academic, 1997).
53. Sheldrick, G. A short history of SHELX. *Acta Crystallogr. A* **64**, 112–122 (2008).
54. Brünger, A. T. et al. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
55. CCP4. The CCP4 Suite: Programs for Protein Crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
56. Zhang, K. Y., Cowtan, K. & Main, P. Combining constraints for electron-density modification. *Methods Enzymol.* **277**, 53–64 (1997).
57. Emsley, P., Lohkamp, B., Scott, W. & Cowtan, K. Features and Development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
58. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
59. Schröder, G. F., Levitt, M. & Brünger, A. T. Super-resolution biomolecular crystallography with low-resolution data. *Nature* **464**, 1218–1222 (2010).
60. Strong, M. et al. Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from *Mycobacterium tuberculosis*. *Proc. Natl Acad. Sci. USA* **103**, 8060–8065 (2006).
61. Zheng, G., Lu, X. J. & Olson, W. K. Web 3DNA—a web server for the analysis, reconstruction, and visualization of three-dimensional nucleic-acid structures. *Nucleic Acids Res.* **37**, W240–W246 (2009).
62. Lavoie, B. D. & Chaconas, G. Site-specific HU binding in the Mu transpososome: conversion of a sequence-independent DNA-binding protein into a chemical nuclease. *Genes Dev.* **7**, 2510–2519 (1993).
63. Swinger, K. K. & Rice, P. A. IHF and HU: flexible architects of bent DNA. *Curr. Opin. Struct. Biol.* **14**, 28–35 (2004).

A primordial origin for misalignments between stellar spin axes and planetary orbits

Konstantin Batygin^{1,2}

The existence of gaseous giant planets whose orbits lie close to their host stars ('hot Jupiters') can largely be accounted for by planetary migration associated with viscous evolution of proto-planetary nebulae¹. Recently, observations of the Rossiter–McLaughlin effect² during planetary transits have revealed that a considerable fraction of hot Jupiters are on orbits that are misaligned with respect to the spin axes of their host stars³. This observation has cast doubt on the importance of disk-driven migration as a mechanism for producing hot Jupiters. Here I show that misaligned orbits can be a natural consequence of disk migration in binary systems whose orbital plane is uncorrelated with the spin axes of the individual stars^{4–6}. The gravitational torques arising from the dynamical evolution of idealized proto-planetary disks under perturbations from massive distant bodies act to misalign the orbital planes of the disks relative to the spin poles of their host stars. As a result, I suggest that in the absence of strong coupling between the angular momentum of the disk and that of the host star, or of sufficient dissipation that acts to realign the stellar spin axis and the planetary orbits, the fraction of planetary systems (including systems of 'hot Neptunes' and 'super-Earths') whose angular momentum vectors are misaligned with respect to their host stars will be commensurate with the rate of primordial stellar multiplicity.

The obliquities (angles between the planetary orbits and the stellar spins) of detected planetary orbits range from almost perfectly aligned prograde to almost perfectly aligned retrograde systems⁷. Previously, the misalignment between planetary orbits and stellar spin axes had been attributed to post-nebular multi-body interactions. Most notably, Kozai cycles with tidal friction^{8–10}, planet–planet scattering^{11,12}, and chaotic secular excursions¹³ have been invoked as a means of producing misaligned planets. These mechanisms are probably responsible for a few specific examples (for example, the extreme eccentricity of HD80606b is almost certainly due to Kozai resonance with the stellar companion HD80607⁸). However, it is unlikely that they can explain misaligned hot Jupiters as a population. For instance, the Kozai mechanism can be stifled by forced apsidal precession in multi-planet systems¹³. Likewise, within the context of planet–planet scattering and secular chaos, the allowed parameter range is limited, because the production of close-in orbits requires the timescale for tidal capture to be considerably shorter than that for eccentricity growth¹², while demanding the associated tidal heating to be small enough not to over-inflate the planet beyond its Roche lobe¹⁴. Additionally, the observed presence of mean-motion resonances among giant planets on wide orbits (which rely on smooth, convergent migration to congregate¹⁵) provides further motivation for the development of a unified model for disk migration that is capable of producing misaligned orbits.

The dynamics of self-gravitating proto-planetary disks under external perturbations can be extremely complex, making precise quantitative modelling computationally unfeasible. Consequently, here I shall concentrate on characterization of the qualitative physical behaviour of the system and use classical perturbation methods to obtain a solution. In the spirit of secular theory¹⁶, I model the proto-planetary disk as a

series of initially planar, circular, concentric, massive wires that interact gravitationally. Our model is based on the Gaussian averaging method^{17,18} and the gravitational potential is softened to partially account for the discrete representation of the disk. The effects of dissipative fluid forces within the disk are neglected. The perturbing body is also modelled as a massive ring, but is eccentric ($e' = 0.5$) and inclined with respect to the disk by an inclination i' . A description of the model and its inherent assumptions is presented in ref. 19, and the details of our implementation are stated in the Supplementary Information.

A self-gravitating disk will preserve an untwisted structure and act as a rigid body, provided that the characteristic timescale of the external perturbation greatly exceeds that of the disk's self-interaction²⁰. Mathematically, this amounts to a statement of adiabatic invariance of the phase-space area occupied by a single secular cycle within the disk²¹. If this condition is satisfied, the external perturber's sole effect is to induce a recession (that is, a retrograde drift) of the ascending node of the disk, as defined by the plane of the stellar orbit. The embedded planetary orbit will also adiabatically follow the disk.

In the reference frame of the host star, the nodal recession of the disk will appear as a cyclic excitation of inclination between the disk and the stellar spin axis (see Fig. 1), provided that the host star's angular momentum vector does not adiabatically trail the disk. For this to hold true, the characteristic interaction timescale between the disk and the

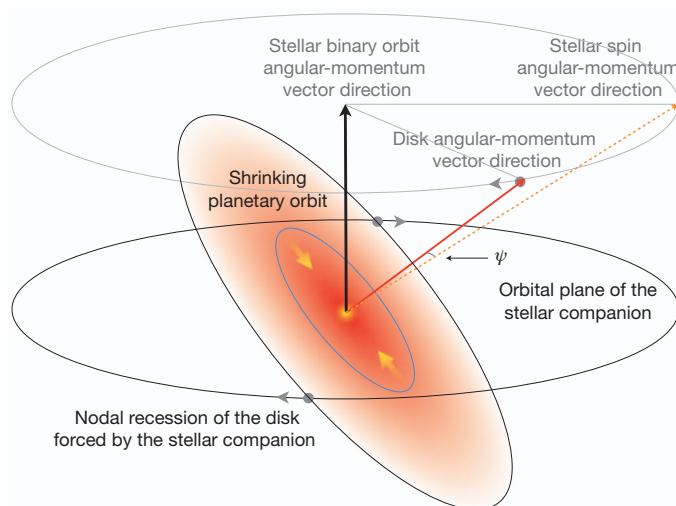


Figure 1 | Geometrical set-up of the problem. This figure depicts a schematic representation of the production of misaligned close-in planets through disk-driven migration in binary systems. The adiabatic response of a self-gravitating disk to long-term perturbations by a stellar companion is the recession of its ascending node, as defined by the orbital plane of the stellar companion. The recession of the disk's angular momentum vector about the stellar binary orbital angular momentum vector appears to be an excitation of the misalignment angle ψ between the stellar spin axis and the disk in the star's reference frame.

¹Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California 91125, USA. ²Institute for Theory and Computation, Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA.

stellar spin-axis, $\mathcal{T}_{\text{star}}$, must exceed the disk's nodal recession timescale, $\mathcal{T}_{\text{disk}}$, by a considerable amount (that is, angular momentum coupling between the disk and the host star must be non-adiabatic)⁹. The former can be estimated by modelling the stellar rotational bulge as an inertially equivalent orbiting ring, effectively reducing the characteristic interaction timescale to the forced nodal recession period of the ring.

Observations suggest that rotational periods of T Tauri pre-main sequence stars, whose masses exceed $M > 0.25 M_{\odot}$, where M_{\odot} is the mass of the Sun, form a bimodal distribution where fast and slow rotators are centred around 2 days and 8 days respectively, with a preference for slow rotation at higher masses²². Thus, for typical pre-main-sequence stars, we obtain $\mathcal{T}_{\text{star}} \approx 10$ Myr and $\mathcal{T}_{\text{star}} \approx 0.3$ Myr for slow and fast rotators, respectively (see Supplementary Information for details). As will be shown below, this suggests that the adiabatic trailing of the stellar spin axes will only prevent excitation of mutual misalignment for fast rotators. Furthermore, the transfer of angular momentum from the disk to the star will probably be unimportant in mature disks because of low accretion rates²³.

In addition to avoiding the adiabatic trailing of the host stars, the prominence of the mechanism described here is determined by the abundance and longevity of wide stellar binary systems in star-formation environments, because the misalignment angle, ψ , becomes fixed when the binary companion is stripped away or when the proto-planetary disk dissipates (although a long-lived binary companion can act to misalign a mature planetary system with its host star²⁴). It is tempting to estimate the frequency with which significant misalignments are attained by this mechanism via population synthesis, but the enormous range and vast observational uncertainties in the input parameters would render such a calculation of little practical use. In particular, although observations of the Taurus-Auriga star-forming region²⁵ suggest that the orbital distribution of young solar-type binaries is roughly log-flat with an overall binary fraction of around 40%, it is noteworthy that the process of wide binary formation also appears to exhibit environmental dependence⁵. Simultaneously, the rate at which wide binaries get disrupted in birth clusters depends sensitively on the local densities within the clusters⁶, which remain observationally elusive, because the majority of stars are born in aggregates that dissolve quickly (on timescales of a few tens of millions of years or less)²⁶. Nevertheless, the above conditions probably imply that the timescale for excitation of significant misalignment should be considerably less than 10 Myr or so.

In systems where self-gravity is strong enough to maintain the effective rigidity of the disk, fast circulation of the disk's argument of perihelion also ensures adiabatic eccentricity dynamics. Specifically, this means that the disk will not develop significant eccentricity, and the Kozai resonance within the individual annuli will be suppressed²⁰ (see Supplementary Information). Additionally, recall that here I am ignoring dissipative effects that would generally act to circularize the disk and maintain its rigidity. Assuming that the angular momentum of the stellar binary greatly exceeds that of the proto-planetary disk, the maximal inclination that can be excited between the stellar rotation axis and the disk's orbital plane is approximately equal to twice the inclination of the proto-planetary disk with respect to the binary orbit: $\psi_{\text{max}} \approx 2i'$. As a result, retrograde planetary orbits can be naturally achieved, provided that the inclination of the stellar companion exceeds 45° . Figure 2 shows the time evolution of two such examples.

Although the Gaussian averaging model used above yields a rigorous representation of the secular evolution of the system, it is also computationally expensive. Fortunately, similar results can be obtained with a modified (arbitrary i' , e') Laplace-Lagrange analytical theory¹⁸ (see Supplementary Information), providing an avenue for efficient mapping of parameter space. We have quantified the precession timescale for a range of stellar companion masses as well as binary separations. The results are summarized in Fig. 3. This calculation highlights an effective equivalence between distant massive perturbers and lower-mass perturbers with smaller semi-major axes, because the recession

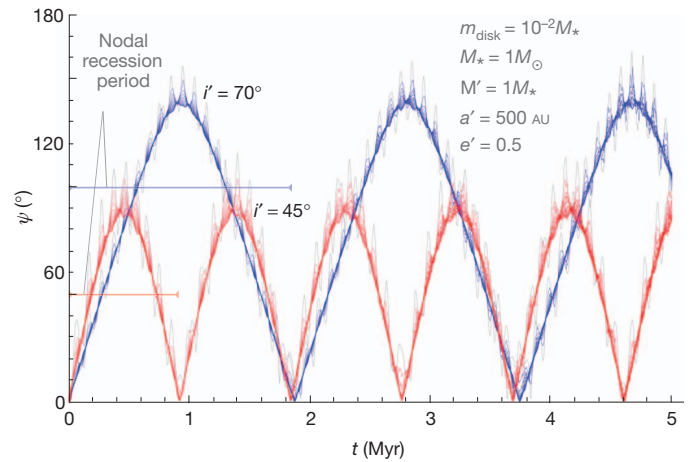


Figure 2 | Excitation of disk–star misalignment. The time evolution of the misalignment angle ψ is shown. The nebula I consider, with disk mass $m_{\text{disk}} = 10^{-2} M_{\odot} \approx 10 M_{\text{Jup}}$, has a surface density profile of the form $\Sigma \propto r^{-1}$, extends between $a_{\text{in}} = 1$ astronomical units (AU) and $a_{\text{out}} = 50$ AU, and orbits a host star with mass $M_{\star} = 1 M_{\odot}$. The plotted curves represent the dynamical states of the disk annuli, where every other ring is plotted. The grey curves depict the innermost annuli. The close proximity of the rings to each other demonstrates that the disk remains locally unwarped and acts as a rigid body to a good approximation. The extent of rigidity is largely controlled by the disk mass, with heavier disks remaining closer to their mid-plane. The host star is assumed to be a slow rotator and its spin pole is held fixed for simplicity. The perturbing $M' = 1 M_{\star}$ binary companion lies $a' = 500$ AU away, has an eccentricity of $e' = 0.5$, and is inclined with respect to the disk by $i' = 45^\circ$ (red) and by $i' = 70^\circ$ (blue). Throughout the duration of the integration, the annuli of the disk never attain significant eccentricity ($e \lesssim 0.1$; see Supplementary Information for an in-depth discussion). This calculation was performed using a conservative softened Gaussian averaging model (with $N_{\text{rings}} = 31$; see Supplementary Information), and thus contains no restrictions on the secular dynamics of the system, but ignores dissipative forces of the gas. The nodal recession periods characteristic of this set-up are $\mathcal{T}_{\text{disk}} \approx 0.9$ Myr ($i' = 45^\circ$) and $\mathcal{T}_{\text{disk}} \approx 1.8$ Myr ($i' = 70^\circ$). The results in this figure can be translated to other system parameters by noting that the maximum misalignment attainable by the disk is roughly twice the disk/binary orbit inclination and that the recession period scales approximately as $\mathcal{T}_{\text{disk}} \propto a'^3 / (M' \cos i' (1 + 3e'^2/2))$ (see also Fig. 3).

period of the disk scales approximately as $\mathcal{T}_{\text{disk}} \propto a'^3 / M'$, where a' is the binary star's semi-major axis and M' is its mass. Thus, the precessional effect of a $M' = 1 M_{\odot}$ star orbiting at $a' = 10^3$ AU is equivalent to the precessional effect that arises from the protoplanetary disk and its host star orbiting an approximately $10^5 M_{\odot}$ star cluster at $a' = 0.25$ pc. I note that bound companions are not necessarily required for production of oblate disks, given that impulsive perturbations from passing stars in the birth cluster will cause the inclination of the disk to execute a random walk and in some cases can excite significant misalignment²⁷. Collectively, this explanation places the approximately 7° misalignment between the Sun's spin axis and the Solar System's invariable plane into a more general, extrasolar context. However, the process of planet formation in perturbed, possibly warped, disks certainly deserves further study.

Given the diverse nature of the environments in which planetary systems may form, one would expect a wide range of characteristic precession timescales for proto-planetary disks. Although this does not necessarily imply an isotropic distribution of spin-orbit angles, it is quite possible that hot Jupiters that emerged from protoplanetary disks in multi-stellar systems already resided on misaligned orbits at the time of nebular dispersion. If this is true, small spin-orbit angles must be in large part either a result of adiabatic trailing of the host stars or dissipative re-alignment of the system. Such a scenario appears to be supported by observations: hot Jupiters orbiting hot, massive stars tend to be misaligned while hot Jupiters orbiting less-massive cooler stars

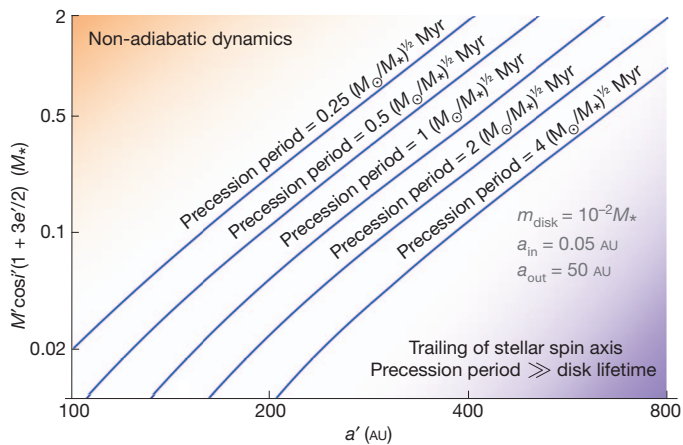


Figure 3 | Timescales for excitation of spin-orbit misalignment. The characteristic nodal recession period of a disk with $m_{\text{disk}} = 10^{-2} M_*$ is shown as a function of binary mass and orbital properties. The disk considered has an outer edge at $a_{\text{out}} = 50$ AU. Increasing a_{out} will result in an approximately linear increase of the precession frequency. The period is expressed as a scaling law in the mass of the host star. In the region of parameter space where the precession period greatly exceeds the disk lifetime, only small misalignment angles between the disk and the host star can be excited. In principle, however, if the stellar companion does not get stripped away, the ascending node of the invariable plane of the formed planetary system can also recess. However, the degree to which this can affect a planet on a close-in orbit is sensitive to the particular architecture of the system. In the region of the parameter space where dynamics ceases to be adiabatic, misalignment is certainly attainable, but more quantitatively precise (magnetohydrodynamical) modelling is required for its characterization. Finally, at the extreme high-mass/small-orbital-separation end of parameter space, one could envision a scenario where a newly formed disk becomes severely twisted and eventually gets disrupted as a result of strong external perturbations. Driven by viscous dissipation, however, such a structure would probably re-collapse into a new protoplanetary disk, whose orbital plane will be close to the Laplace plane of the stellar binary orbit.

tend to have small spin-orbit angles³. This transition has been attributed to tidal re-alignment of cooler stars owing to the increased size of their convective zones²⁸ and thus appears to be in good agreement with our model.

The consistency arguments presented above indicate that disk-driven migration in binary systems is a favourable origin of misaligned hot Jupiters. This model can be tested as follows. Although multiple explanations exist for the origins of hot Jupiters, short-period multi-planet systems, which are typically less massive, have almost certainly undergone disk-driven migration. Theoretically, disk-driven migration tends to maintain coplanarity and near-resonances in the planetary systems²⁹, both of which have been spectacularly confirmed by the Kepler mission³⁰. Within the context of the model proposed here, proto-planetary disks become misaligned with the spin axes of their host stars irrespectively of the masses and orbital radii of the newly formed planets. I predict that future observations of the Rossiter–McLaughlin effect will reveal that systems of close-in coplanar sub-giant planets can be misaligned with respect to the spin axes of their host stars in the absence of significant dissipative processes. Similarly, the presence of distant, nearly circular resonant planet pairs in systems that host hot Jupiters on oblique orbits would also point to disk-torque as the likely mechanism for the origin of spin-orbit misalignment. High-precision radial velocity monitoring of transiting systems should directly test this prediction in the near future.

Received 2 November 2011; accepted 24 August 2012.

1. Lin, D. N. C., Bodenheimer, P. & Richardson, D. C. Orbital migration of the planetary companion of 51 Pegasi to its present location. *Nature* **380**, 606–607 (1996).

2. McLaughlin, D. B. Some results of a spectrographic study of the Algol system. *Astrophys. J.* **60**, 22–31 (1924).
3. Winn, J. N., Fabrycky, D., Albrecht, S. & Johnson, J. A. Hot stars with hot Jupiters have high obliquities. *Astrophys. J.* **718**, L145–L149 (2010).
4. Ghez, A. M., Neugebauer, G. & Matthews, K. The multiplicity of T Tauri stars in the star forming regions Taurus-Auriga and Ophiuchus-Scorpius: a 2.2 micron speckle imaging survey. *Astron. J.* **106**, 2005–2023 (1993).
5. Kraus, A. L., Ireland, M. J., Martinache, F. & Hillenbrand, L. A. Mapping the shores of the brown dwarf desert. II. Multiple star formation in Taurus-Auriga. *Astrophys. J.* **731**, 8 (2011).
6. Marks, M. & Kroupa, P. Inverse dynamical population synthesis: constraining the initial conditions of young stellar clusters by studying their binary populations. *Astron. Astrophys.* **543**, A8 (2012).
7. Hébrard, G. 20 colleagues The retrograde orbit of the HAT-P-6b exoplanet. *Astron. Astrophys.* **527**, L11 (2011).
8. Wu, Y. & Murray, N. Planet migration and binary companions: the case of HD 80606b. *Astrophys. J.* **589**, 605–614 (2003).
9. Fabrycky, D. & Tremaine, S. Shrinking binary and planetary orbits by Kozai cycles with tidal friction. *Astrophys. J.* **669**, 1298–1315 (2007).
10. Naoz, S., Farr, W. M., Lithwick, Y., Rasio, F. A. & Teyssandier, J. Hot Jupiters from secular planet–planet interactions. *Nature* **473**, 187–189 (2011).
11. Ford, E. B. & Rasio, F. A. Origins of eccentric extrasolar planets: testing the planet–planet scattering model. *Astrophys. J.* **686**, 621–636 (2008).
12. Nagasawa, M., Ida, S. & Bessho, T. Formation of hot planets by a combination of planet scattering, tidal circularization, and the Kozai mechanism. *Astrophys. J.* **678**, 498–508 (2008).
13. Wu, Y. & Lithwick, Y. Secular chaos and the production of hot Jupiters. *Astrophys. J.* **735**, 109 (2011).
14. Guillot, J., Ramirez-Ruiz, E. & Lin, D. Consequences of the ejection and disruption of giant planets. *Astrophys. J.* **732**, 74 (2011).
15. Morbidelli, A. & Crida, A. The dynamics of Jupiter and Saturn in the gaseous protoplanetary disk. *Icarus* **191**, 158–171 (2007).
16. Marquis de Laplace, P.-S. *Traité de Mécanique Céleste*, 1 and 2 Vol. 1, Ch. VII, 569–634, <http://ia600303.us.archive.org/19/items/mcaniquecles01laprich/mcaniquecles01laprich.pdf> (Hillard, Gray, Little and Wilkins, 1799).
17. Gauss, C. F. *Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientium* (eds Perthes, F. & Besser, I. H.) (Sumtibus, 1809); in *Werke* Vol. 3, 331–355, <http://resolver.sub.uni-goettingen.de/purl?PPN235999628> (Dieterich, Gottinger, 1866).
18. Murray, C. D. & Dermott, S. F. *Solar System Dynamics* Ch. 7 (Cambridge University Press, 1999).
19. Tórama, J. R., Tremaine, S. & Kazandjian, M. V. Gauss's method for secular dynamics, softened. *Mon. Not. R. Astron. Soc.* **394**, 1085–1108 (2009).
20. Batygin, K., Morbidelli, A. & Tsiganis, K. Formation and evolution of planetary systems in presence of highly inclined stellar perturbers. *Astron. Astrophys.* **533**, A7 (2011).
21. Morbidelli, A. *Modern Celestial Mechanics: Aspects of Solar System Dynamics* Chs 4 and 8 (Taylor and Francis, 2002).
22. Herbst, W., Bailer-Jones, C. A. L., Mundt, R., Meisenheimer, K. & Wackermann, R. Stellar rotation and variability in the Orion nebula cluster. *Astron. Astrophys.* **396**, 513–532 (2002).
23. Calvet, N. et al. Disk evolution in the Orion OB1 association. *Astron. J.* **129**, 935–946 (2005).
24. Kaib, N. A., Raymond, S. N. & Duncan, M. J. 55 Cancri: a coplanar planetary system that is likely misaligned with its star. *Astrophys. J.* **742**, L24 (2011).
25. Kraus, A. L. & Hillenbrand, L. A. The role of mass and environment in multiple-star formation: a 2MASS survey of wide multiplicity in three young associations. *Astrophys. J.* **662**, 413–430 (2007).
26. Adams, F. C. The birth environment of the Solar System. *Annu. Rev. Astron. Astrophys.* **48**, 47–85 (2010).
27. Bate, M. R., Lodato, G. & Pringle, J. E. Chaotic star formation and the alignment of stellar rotation with disc and planetary orbital axes. *Mon. Not. R. Astron. Soc.* **401**, 1505–1513 (2010).
28. Lai, D. Tidal dissipation in planet-hosting stars: damping of spin-orbit misalignment and survival of hot Jupiters. *Mon. Not. R. Astron. Soc.* **423**, 486–492 (2011).
29. Terquem, C. & Papaloizou, J. C. B. Migration and the formation of systems of hot super-Earths and Neptunes. *Astrophys. J.* **654**, 1110–1120 (2007).
30. Lissauer, J. J. et al. Architecture and dynamics of Kepler's candidate multiple transiting planet systems. *Astrophys. J. Suppl. Ser.* **197**, 8 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements I thank A. Morbidelli, P. Golreich, H. Knutson, S. Tremaine, J. Winn and F. Adams for numerous conversations and D. Stevenson and G. Laughlin for carefully reading the manuscript. I am greatly indebted to M. Kazandjian and J. Tórama for providing me with the softened analytical Gaussian averaging algorithm used in this work and help with implementation. Finally, I am grateful to D. Fabrycky for careful examination of the paper and numerous suggestions, which resulted in a substantial improvement of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The author declares no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.B. (kbatygin@gps.caltech.edu).

Quantum-dot spin–photon entanglement via frequency downconversion to telecom wavelength

Kristiaan De Greve^{1†}, Leo Yu^{1*}, Peter L. McMahon^{1*}, Jason S. Pelc^{1*}, Chandra M. Natarajan^{1,2}, Na Young Kim¹, Eisuke Abe^{1,3}, Sebastian Maier⁴, Christian Schneider⁴, Martin Kamp⁴, Sven Höfling^{1,4}, Robert H. Hadfield², Alfred Forchel⁴, M. M. Fejer¹ & Yoshihisa Yamamoto^{1,3}

Long-distance quantum teleportation and quantum repeater technologies require entanglement between a single matter quantum bit (qubit) and a telecommunications (telecom)-wavelength photonic qubit^{1–5}. Electron spins in III–V semiconductor quantum dots are among the matter qubits that allow for the fastest spin manipulation^{6,7} and photon emission^{8,9}, but entanglement between a single quantum-dot spin qubit and a flying (propagating) photonic qubit has yet to be demonstrated. Moreover, many quantum dots emit single photons at visible to near-infrared wavelengths, where silica fibre losses are so high that long-distance quantum communication protocols become difficult to implement¹⁰. Here we demonstrate entanglement between an InAs quantum-dot electron spin qubit and a photonic qubit, by frequency downconversion of a spontaneously emitted photon from a singly charged quantum dot to a wavelength of 1,560 nanometres. The use of sub-10-picosecond pulses at a wavelength of 2.2 micrometres in the frequency downconversion process provides the necessary quantum erasure to eliminate which-path information in the photon energy. Together with previously demonstrated indistinguishable single-photon emission at high repetition rates^{11,12}, the present technique advances the III–V semiconductor quantum-dot spin system as a promising platform for long-distance quantum communication.

A quantum communication network¹ will consist of stationary matter qubits and flying photonic qubits. The two key technologies for quantum communication networks are quantum teleportation from a photonic qubit to a matter qubit or vice versa, and a quantum repeater to create and store the entangled states of remote matter qubits. These two core technologies rely on the ability to create entanglement between a matter qubit and a photonic qubit over long distances, followed by a Bell state measurement that transforms the matter qubit–photonic qubit entanglement into the teleportation of a given unknown quantum state, or into matter–matter entanglement^{13–18}. Matter–photon and matter–matter entanglement generation have been shown in several ionic^{4,16,17} and atomic^{2,5,18} systems, while matter–photon entanglement was recently observed in nitrogen-vacancy diamond colour centres³. However, these systems suffer from low photon collection efficiencies and relatively long optical recombination times, which can only partially be overcome by cavity–quantum electrodynamic (cavity–QED) solutions¹. In addition, none of them use telecom-wavelength (1.5 μm) photons that would allow long-distance entanglement distribution. Combining a solid-state monolithic cavity–QED system with a fast quantum emitter^{19–21} could solve the low photon yields. In the InAs quantum-dot system, subnanosecond optical recombination times have previously led to the demonstration of high single-photon rates when the quantum dots are embedded in high-quality optical cavities^{8,9,20,21}. Moreover, spin control of charged InAs quantum dots was shown to be feasible on picosecond timescales

while preserving spin coherence^{6,7,22}, making it a promising candidate as a quantum network technology. However, spin–photon entanglement has yet to be established in this system, and its emission wavelength does not match the low-loss wavelength range of silica fibres. In this work, we address both these challenges with an ultrafast frequency downconversion technique based on a periodically poled lithium niobate (PPLN) waveguide device²³. Its timing resolution enables demonstration of entanglement between a single InAs quantum-dot spin qubit and a photonic qubit at 910 nm, while the target wavelength (1,560 nm) would permit long-distance quantum communication.

The physical system used in our experiment is presented in Fig. 1a–d. It consists of a single electron-doped InAs quantum dot, embedded in a low- Q microcavity (see also Methods). With an external magnetic field oriented in the Voigt geometry (perpendicular to the growth direction/optical axis), two Λ -systems are formed, where an optically active, excited state is coupled to each of two ground states. The excited states are so-called trion states, as they consist of three particles: two electrons, paired into a singlet state, and an unpaired hole ($|\uparrow\downarrow\uparrow\rangle$ and $|\uparrow\downarrow\downarrow\rangle$). Both trion states are connected to each of the electron spin states, $|\downarrow\rangle$ and $|\uparrow\rangle$, in a Λ -configuration. These Λ -systems have been extensively studied²⁴, and were previously used for initialization²⁵ and coherent manipulation of the electron spin states^{6,7,22,26}. The optical selection rules are indicated in Fig. 1a, and were verified by polarization-selective photoluminescence (Fig. 1b). In addition, initialization, rotation and measurement of the spin state of the quantum dot are realized, using all-optical techniques as reported previously^{6,7} (see Fig. 1c, d and Methods). By selectively exciting one of the trion states, for example, the $|\uparrow\downarrow\uparrow\rangle$ -state, spontaneous emission decay in a Λ -system leads to entanglement between the emitted photon and the electron spin^{3,4}:

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|\uparrow\rangle \otimes |iH; \omega + \delta\omega\rangle + |\downarrow\rangle \otimes |V; \omega\rangle) \quad (1)$$

Here $\delta\omega$ refers to the difference in energy (colour) of the emitted photons, determined by the electron Zeeman energy, and H and V to their polarization (see Fig. 1c); ω refers to the frequency (wavelength) of the V -polarized photon (910.10 nm). Previous coherent spin manipulation experiments^{6,7} relied on a relatively large Zeeman splitting $\delta\omega$ between the respective electron spin states, both for high-fidelity spin initialization and readout, and for ultrafast coherent control of the spin through a combination of stimulated-Raman transitions and Larmor precession. In the present experiment, this energy separation is set at about $2\pi \times 17.6$ GHz (external magnetic field $B = 3$ T, Fig. 1d), leaving room for which-path information to leak out into the environment via the photon energy, which makes verification of entanglement between the polarization of the photon and the electron spin challenging. Upon detection of an emitted photon in

¹E. L. Ginzton Laboratory, Stanford University, Stanford, California 94305, USA. ²Scottish Universities Physics Alliance and School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh EH14 4AS, UK. ³National Institute of Informatics, Hitotsubashi 2-1-2, Chiyoda-ku, Tokyo 101-8403, Japan. ⁴Technische Physik, Physikalisches Institut, Wilhelm Conrad Röntgen Research Center for Complex Material Systems, Universität Würzburg, Am Hubland, D-97074 Würzburg, Germany. [†]Present address: Department of Physics, Harvard University, 17 Oxford Street, Cambridge, Massachusetts 02138, USA.

*These authors contributed equally to this work.

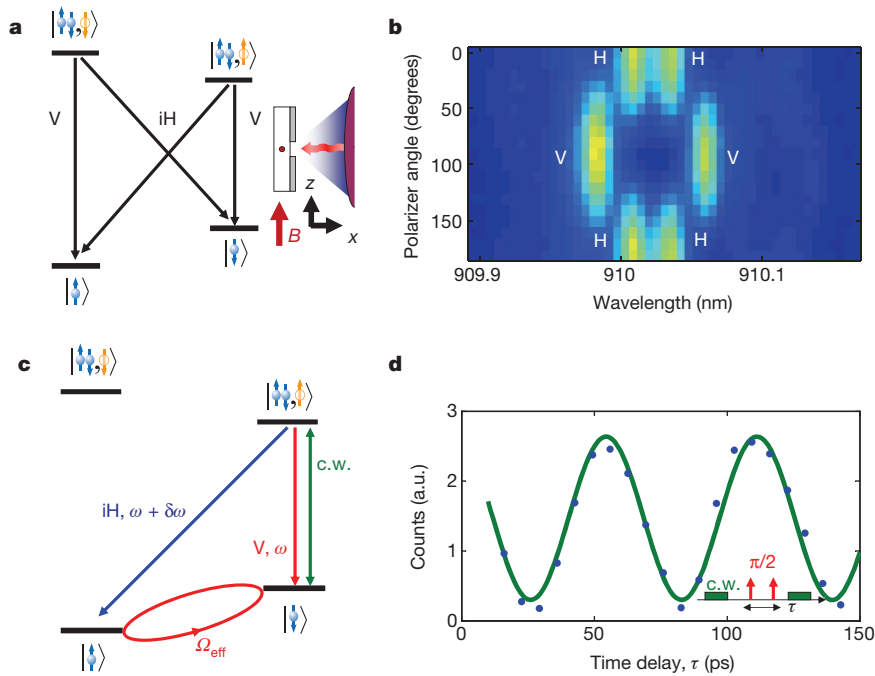


Figure 1 | Level structure of quantum dot and spin manipulation. **a**, Level structure of an electron-doped quantum dot, with the magnetic field in-plane (Voigt geometry, boxed). V and H refer to linear polarizations, either perpendicular (V) or parallel (H) to the magnetic field. Blue and orange arrows refer to the electron and (trion-) hole states, respectively. **b**, Verification (by polarization-selective magnetophotoluminescence) of the polarization selection rules of the studied electron-doped quantum dot. **c**, Level structure and optical spin manipulation scheme used. $\delta\omega$, Larmor precession frequency ($2\pi \times 17.6$ GHz for $B = 3$ T); ω , frequency of a V-polarized photon (wavelength 910.10 nm for $B = 3$ T); c.w., narrowband, continuous wave laser used for initialization and readout; Ω_{eff} , effective spin Rabi frequency resulting from manipulation by detuned ($\Delta = 300$ GHz), circularly polarized optical pulses^{6,7}. **d**, Ramsey interference experiment, demonstrating coherent control of the electron spin qubit. Blue, raw data, for $\pi/2$ - $\pi/2$ interference; green, least squares fit. For a 3 T magnetic field, the Larmor precession frequency $\delta\omega = 2\pi \times 17.6$ GHz. Inset, pulse scheme used.

the rotated polarization basis (say, $|\sigma^+\rangle$ -photon) at time t_1 , the resulting spin state evolves as follows:

$$|\psi_{\text{spin},\sigma^+}(t)\rangle = \frac{1}{\sqrt{2}} \left(e^{i\delta\omega(t-t_1)} |\uparrow\rangle - |\downarrow\rangle \right) \quad (2)$$

Owing to the large Zeeman splitting, a small uncertainty Δt_1 in the photon's arrival time can result in a large uncertainty in the phase of the spin state ($\delta\omega \times \Delta t_1$) and therefore a reduction in the entanglement visibility. Lowering the magnetic field further reduces the fidelity of the single-photon-based spin readout⁶ and is therefore not a viable option in our experiment. Previous experiments in ionic⁴ and NV-diamond systems³ relied on the speed (timing jitter) of commercial single-photon detectors to fix the phase of the coherent spin precession, or, equivalently, to act as a quantum eraser²⁷ of the relatively small energy difference between different branches of the Λ -system (Supplementary Information). Such high bandwidth/fast detection techniques have also been employed to increase visibility in photon-photon interference experiments¹². However, the much larger energy difference in our system requires a more advanced quantum eraser technique.

In order to tackle the large energy difference, we use a frequency downconversion technique that converts a single 910-nm photon into a single 1,560-nm photon with sub-8-ps timing resolution; this extreme timing resolution is sufficient to erase the which-path frequency information. The technique consists of mixing the single 910-nm photon with a few-picosecond pump light pulse at 2.2 μm in a PPLN waveguide device, producing a 1,560-nm photon via the process of difference-frequency generation (Methods). After narrowband filtering at 1,560 nm, the frequency-downconverted photon can be detected on a superconducting nanowire single-photon detector²⁸ (SNSPD) conditional on the exact overlap of the single 910-nm photon and the 2.2- μm pump pulse. The detection of the 1,560-nm photon therefore heralds a time-accurate measurement of the 910-nm photon: only if the 910-nm photon entered the PPLN waveguide exactly at the same time as the 2.2- μm pulse can a 1,560-nm photon be generated.

The system diagram of the conversion set-up is indicated in Fig. 2a. Using cross-correlation with a bright, 3-ps 910-nm pulse (Fig. 2b), we can infer a timing resolution of 8 ps or better for the PPLN waveguide device (Methods). As 8 ps correspond to an effective bandwidth of more

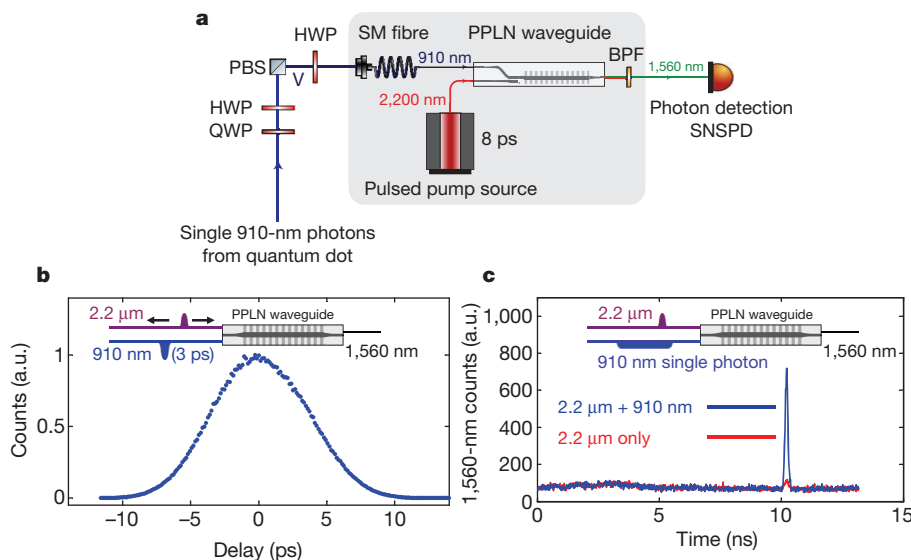


Figure 2 | Ultrafast conversion to 1,560 nm. **a**, Schematic overview of the conversion technique used (Methods). The 910-nm photon polarization is measured, with high timing accuracy, by a combination of a polarization analysing stage, a PPLN downconverter and an SNSPD. PBS, polarizing beamsplitter; HWP, half-wave plate; QWP, quarter-wave plate; SM, single-mode; BPF, 1,560 nm bandpass filter. **b**, Timing resolution of the 1,560-nm conversion technique, measured by cross-correlating the 2.2- μm conversion pulse with a classical, 3-ps pulse at 910 nm (inset). From these data, we can infer a sub-8-ps resolution for the arrival time of a single photon from the quantum dot. **c**, Performance of the conversion technique at single-photon levels, measured using an SNSPD. For single, 910-nm photons at the input (blue trace), the residual noise (red trace) can be seen to be well below the single-photon level (Methods). Inset, cross-correlation set-up.

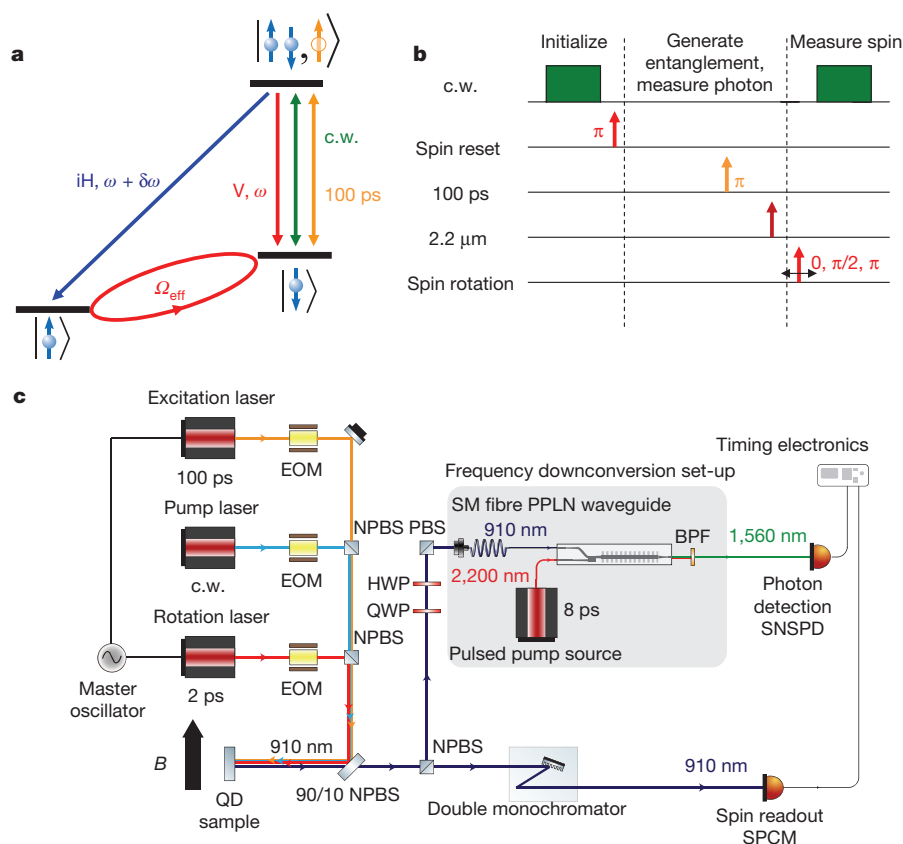


Figure 3 | Quantum-dot manipulation scheme for spin-photon entanglement verification. **a**, Schematic overview of the quantum-dot manipulation techniques used in the experiment. ‘100 ps’ indicates the resonant, 100-ps laser pulse used to excite the system into the $|\uparrow\downarrow\rangle$ -state; ‘c.w.’ indicates the continuous wave (c.w.) laser used for initialization into the $|\uparrow\rangle$ -state, and for readout of the $|\downarrow\rangle$ -state. $\delta\omega$, Spin Larmor precession frequency; Ω_{eff} , effective spin Rabi frequency, resulting from manipulation by 3-ps, detuned optical pulses^{6,7}. For the classical correlation measurements, manipulation of the spin using the H-branch was used as well, whereas the entanglement result was obtained using only V-branch pumping. **b**, Timing

diagram of the pulse sequence used in the experiment. Note the three cycles used in the experiment: initialization through optical pumping and spin rotation, generation of entanglement using 100-ps laser pulses and spontaneous emission decay, followed by ultrafast photon conversion to a telecom wavelength, and spin measurement through a combination of spin rotation (measurement basis change), optical pumping and single-photon detection. The cycle time of a single shot of the experiment was chosen to be either 39 or 52 ns. **c**, Schematic diagram of the set-up used in the experiment. EOM, electro-optic modulator; NPBS, non-polarizing beamsplitter. See text for other components.

than 100 GHz, this is more than sufficient to erase the information inherent in the Zeeman energy of 17.6 GHz at 3 T (Supplementary Information). In addition, the conversion and filtering technique results in almost noise-free signal transduction to the telecom band, as shown in Fig. 2c. We measure the polarization state of 910-nm photons using a polarization analysing stage consisting of a quarter-wave plate and a half-wave plate and a polarizer, followed by a single downconversion set-up that provides the necessary timing accuracy (filtering) in order to verify entanglement at 910 nm. The polarization selective operation of PPLN waveguides does not permit a full photonic polarization qubit at 910 nm to be directly downconverted to a polarization qubit at 1,560 nm. However, and as a straightforward extension of the work presented in this Letter, mapping the polarization qubit into a dual-rail qubit²⁹, using two separate time-resolved downconverters, would realize a spin-entangled 1,560-nm qubit, suitable for long-distance quantum communication¹⁰ (Supplementary Information).

We analyse spin-photon entanglement using a combination of the time-resolved conversion technique and previously established spin initialization, manipulation and readout techniques^{6,7}. Using 3-ps-long, 1-nm red-detuned optical pulses in a Ramsey interferometry set-up, we can completely control the state of any arbitrary electron spin superposition, with fidelities around 95% for the particular quantum dot used in this work (Supplementary Information). The spin state is measured using an optical pumping scheme^{6,7}, which is also used to initialize the system into the $|\uparrow\rangle$ -state. Both processes emit a single

910-nm H-polarized photon that can be detected after frequency and polarization filtering from the optical pumping laser^{6,7} (Fig. 3a and Methods). For a 13-ns spin interrogation time, readout and initialization fidelities of $\geq 96\%$ can be inferred from the time-resolved emission decay, mainly limited by residual leakage of the continuous-wave laser used for optical pumping (Supplementary Information). The full control sequence of the optical pulses is indicated in Fig. 3b, and consists of three stages. First, the system is initialized into either the $|\downarrow\rangle$ -state or the $|\uparrow\rangle$ -state by a combination of optical pumping and spin rotation with a π -pulse. Then, a 100-ps optical π -pulse resonant with the $|\downarrow\rangle$ - $|\uparrow\downarrow\rangle$ -transition (or $|\uparrow\rangle$ - $|\uparrow\downarrow\rangle$ -transition) excites to the trion state, after which spontaneous emission occurs into the electron spin states. The difference between the spontaneous emission decay time (600 ps) and the excitation pulse duration (100 ps) allows for temporal filtering of the spontaneously emitted single photons. In addition, for the classical correlation measurements (see below), cross-polarization filtering can be used to further suppress noise from the reflection of the 100-ps excitation laser. The spontaneously emitted single photons are then sent to a polarization analysing stage, after which they are converted to 1,560 nm using the PPLN waveguide, and detected by an SNSPD. The next stage of the experiment consists of the spin analysis, which is performed through a combination of spin rotations (these implement an effective measurement basis change) and the optical pumping/readout cycle. We measure the correlations between the polarization of the 910-nm photon and the spin state through a timing histogram analysis

of the photons detected in the conversion and measurement cycles (Fig. 3c and Methods).

Using our time-resolved downconversion technique, we measure the classical (computational basis) correlations between the spin along the magnetic field axis (z), and the H–V polarization of the photon

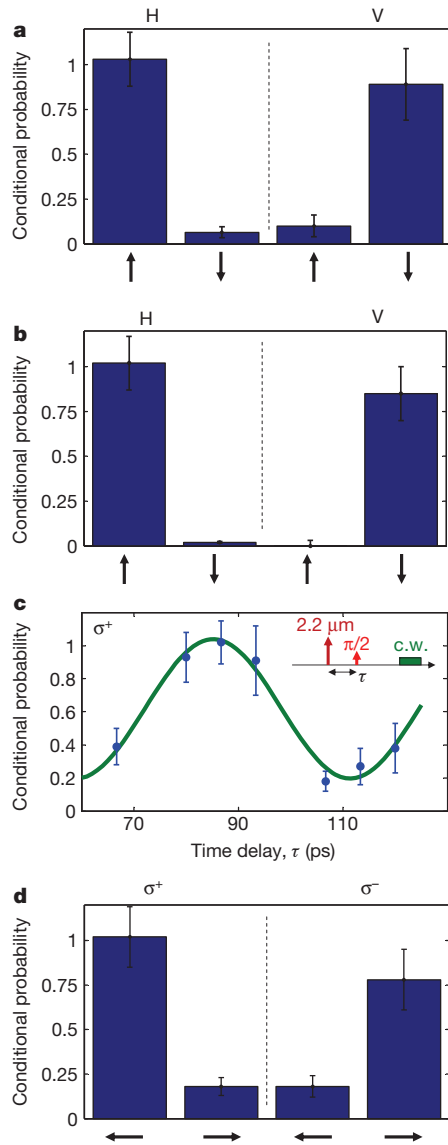


Figure 4 | Spin-photon entanglement verification. **a**, Classical (computational basis) spin-photon correlations, measured through downconversion to 1,560 nm. The black arrows (see also **b**, **d**) refer to the spin orientation in the computational (vertical arrows) or rotated (horizontal arrows) basis. **b**, Classical (computational basis) spin-photon correlations, measured at 910 nm, without downconversion. **c**, Spin-photon entanglement: on time-resolved detection of a $\sigma^{+,-}$ -downconverted photon, the electron spin starts precessing due to the electron Zeeman energy (Larmor precession). By changing the arrival time of the subsequent $\pi/2$ -pulse, this coherent spin precession can be mapped into a Ramsey fringe, where the population in the $|\downarrow\rangle$ -state oscillates as a function of time, and depends on the helicity of the downconverted photon (the Ramsey fringes for opposite helicities are in antiphase). This oscillatory spin signal then results in an oscillatory, periodic coincidence count rate between the downconverted single photon and the spin-measurement photon. Blue, raw data from histogram analysis; green, least-squares, sinusoidal fit to the data. Inset, relative timings used. 2.2 μm indicates the 2.2- μm conversion pump pulse; $\pi/2$ shows the $\pi/2$ spin rotation pulse; c.w. indicates the continuous wave (c.w.), spin measurement laser at 910 nm. **d**, Spin-photon correlations in the rotated bases, demonstrating entanglement. Error bars, ± 1 s.d. (**a–d**), due to Poissonian statistics ($n_{\text{events}} \approx 50\text{--}100$).

(Fig. 4a). For an H-polarized photon, we measure an excellent correlation with the $|\uparrow\rangle$ -spin-state in the subsequent spin measurement cycle. Likewise, for a V-polarized photon, we see a strong correlation with the $|\downarrow\rangle$ -spin-state. We can directly compare these correlations with the ones obtained without ultrafast downconversion, using a commercial single-photon detector (Fig. 4b). The results are in good agreement, and differ only due to small amounts of residual noise in the down-conversion process (Supplementary Information).

Verification of entanglement requires observation of correlations in a rotated basis of the photon polarization and the spin as well. When we measure the photon in the circular polarization basis ($|\sigma^+\rangle$, $|\sigma^-\rangle$), we measure the spin in the basis of $|\rightarrow\rangle = \frac{1}{\sqrt{2}}(|\uparrow\rangle + |\downarrow\rangle)$ and $|\leftarrow\rangle = \frac{1}{\sqrt{2}}(|\uparrow\rangle - |\downarrow\rangle)$. After detection of a $|\sigma^+\rangle$ ($|\sigma^-\rangle$)-downconverted photon at time t_1 , the spin is projected into the $|\leftarrow\rangle$ -state ($|\rightarrow\rangle$), which subsequently evolves in time due to Larmor precession:

$$|\psi_{\text{spin}}(t)\rangle = \frac{1}{\sqrt{2}} \left(e^{i(\delta\omega)(t-t_1)} |\rightarrow\rangle \mp |\leftarrow\rangle \right) \quad (3)$$

Here $\delta\omega$ corresponds to the Zeeman frequency of $2\pi \times 17.6$ GHz. By scanning the arrival time of a $\pi/2$ spin rotation pulse in a Ramsey interferometer^{6,7} (Fig. 4c and Methods) we can trace out this coherent oscillation, and verify entanglement. From the minima and maxima of these coherent oscillations, we can derive, as in Fig. 4d, the photon-spin correlations for $|\sigma^+\rangle$ ($|\sigma^-\rangle$) downconversion. For a particular photon polarization ($|\sigma^+\rangle$, $|\sigma^-\rangle$) and arrival time of the $\pi/2$ spin rotation pulse, a correlation can be measured, as expected. Subsequently changing the arrival time of the $\pi/2$ -pulse by half a Larmor period results in an anticorrelation (see also Fig. 4c). Together, these results demonstrate spin-photon entanglement. We estimate the entanglement fidelity^{3,4} to be 0.8 ± 0.085 (Methods and Supplementary Information), which exceeds the classical limit of 0.5 by more than three standard deviations^{3,4}.

Whereas our entanglement verification technique relies on time-resolved frequency conversion, resulting in filtering out only those 910-nm photons that exactly overlap with the arrival time of the 2.2- μm pump pulse, a more generic downconversion technique with a continuous wave pump laser at 2.2 μm and coincidence detection of two photons from separate quantum-dot sources could be used in future experiments in order to obtain spin-spin entanglement¹⁷ (Supplementary Information).

We have demonstrated high-fidelity spin-photon entanglement in a single InAs quantum dot using an ultrafast downconversion technique to the lowest-loss telecom wavelength (1,560 nm). When combined with ultrafast control of InAs quantum-dot electron spins^{6,7} and fast radiative decay of indistinguishable single photons from such quantum dots embedded in optical cavities^{11,12,20,21}, our results should enable quantum state transfer from a flying qubit to a stationary qubit or vice versa, and entanglement distribution between two remote matter qubits in long-distance quantum networks.

We note that, contemporaneously with this work, another group verified spin-energy entanglement between an InAs quantum-dot electron spin and a spontaneously emitted photon at 966 nm (ref. 30).

METHODS SUMMARY

Quantum dot and optical control. All results are obtained from a single quantum dot, emitting at 910 nm, embedded in a planar microcavity⁷. An external magnetic field in Voigt geometry splits the electron spin and trion states (Fig. 1a). A 0.68 NA aspheric lens focuses the pump and rotation lasers onto the sample. The coherent manipulation techniques are identical to those reported previously⁷, with fidelities (initialization, readout, coherent rotation) around 95% or higher (Supplementary Information). After initialization, the $|\uparrow\downarrow\rangle$ -state is excited by a 100-ps mode-locked-laser pulse, synchronized with the spin-rotation laser. Photoluminescence is collected in a confocal set-up, and split into two branches. One branch is cross-polarized and sent through a double-monochromator onto a single-photon counter (SPCM) for spin-state analysis. The other branch is sent to a polarization analysing stage, after which it is sent to the downconversion set-up.

Downconversion, data acquisition and processing. The 2.2- μm light pulses needed for conversion are generated in a PPLN chip by mixing 3-ps, 911-nm pulses from the spin-rotation laser with narrowband, continuous wave 1,560-nm light. A PPLN waveguide converts 910-nm photons into 1,560-nm photons, conditional on overlap with the 2.2- μm pulses. An SNSPD subsequently detects the 1,560-nm photons. The SPCM and SNSPD signals are combined on a timing analyser, which allows for signal gating in post-processing. The noise in the conversion process is well below the single-photon level (between 4:1 and 10:1 signal-to-noise ratio). The correlation data are obtained from the coincidences between the downconverted single photons and the single photons used for spin detection, through post-processing of the data stream from the timing analyser, and normalized to uncorrelated events (Supplementary Information). The entanglement fidelity analysis follows the same procedure used in ion-trap⁴ and NV-diamond³ spin-photon entanglement experiments, and is estimated to be around 0.8 ± 0.085 , well above the classical limit of 0.5.

Full Methods and any associated references are available in the online version of the paper.

Received 7 June; accepted 12 September 2012.

- Kimble, H. J. The quantum internet. *Nature* **453**, 1023–1030 (2008).
- Ritter, S. *et al.* An elementary quantum network of single atoms in optical cavities. *Nature* **484**, 195–200 (2012).
- Togan, E. *et al.* Quantum entanglement between an optical photon and a solid-state spin qubit. *Nature* **466**, 730–734 (2010).
- Blinov, B. B., Moehring, D. L., Duan, L.-M. & Monroe, C. Observation of entanglement between a single trapped atom and a single photon. *Nature* **428**, 153–157 (2004).
- Wilk, T., Webster, S. C., Kuhn, A. & Rempe, G. Single-atom single-photon quantum interface. *Science* **317**, 488–490 (2007).
- Press, D., Ladd, T. D., Zhang, B. & Yamamoto, Y. Complete quantum control of a single quantum dot spin using ultrafast optical pulses. *Nature* **456**, 218–221 (2008).
- Press, D. *et al.* Ultrafast optical spin echo in a single quantum dot. *Nature Photon.* **4**, 367–370 (2010).
- Pelton, M. *et al.* Efficient source of single photons: a single quantum dot in a micropost microcavity. *Phys. Rev. Lett.* **89**, 233602 (2002).
- Moreau, E. *et al.* A single-mode solid-state source of single photons based on isolated quantum dots in a micropillar. *Physica E* **13**, 418–422 (2002).
- Takesue, H. *et al.* Quantum key distribution over a 40-dB channel loss using superconducting single-photon detectors. *Nature Photon.* **1**, 343–348 (2007).
- Santori, C., Fattal, D., Vuckovic, J., Solomon, G. S. & Yamamoto, Y. Indistinguishable photons from a single-photon device. *Nature* **419**, 594–597 (2002).
- Patel, R. B. *et al.* Two-photon interference of the emission from electrically tunable remote quantum dots. *Nature Photon.* **4**, 632–635 (2010).
- Duan, L.-M., Lukin, M. D., Cirac, J. I. & Zoller, P. Long-distance quantum communication with atomic ensembles and linear optics. *Nature* **414**, 413–418 (2001).
- Briegel, H.-J., Dür, W., Cirac, J. I. & Zoller, P. Quantum repeaters: the role of imperfect local operations in quantum communication. *Phys. Rev. Lett.* **81**, 5932–5935 (1998).
- Yuan, Z.-S. *et al.* Experimental demonstration of a BDCZ quantum repeater node. *Nature* **454**, 1098–1101 (2008).
- Stute, A. *et al.* Tunable ion-photon entanglement in an optical cavity. *Nature* **485**, 482–485 (2012).
- Moehring, D. L. *et al.* Entanglement of single-atom quantum bits at a distance. *Nature* **449**, 68–71 (2007).
- Chou, C. W. *et al.* Measurement-induced entanglement for excitation stored in remote atomic ensembles. *Nature* **438**, 828–832 (2005).
- Faraon, A., Barclay, P. E., Santori, C., Fu, K.-M. C. & Beausoleil, R. G. Resonant enhancement of the zero-phonon emission from a colour centre in a diamond cavity. *Nature Photon.* **5**, 301–305 (2011).
- Michler, P. *et al.* A quantum dot single-photon turnstile device. *Science* **290**, 2282–2285 (2000).
- Santori, C., Pelton, M., Solomon, G., Dale, Y. & Yamamoto, Y. Triggered single photons from a quantum dot. *Phys. Rev. Lett.* **86**, 1502–1505 (2001).
- Kim, D., Carter, S. G., Grelich, A., Bracker, A. S. & Gammon, D. Ultrafast optical control of entanglement between two quantum-dot spins. *Nature Phys.* **7**, 223–229 (2011).
- Pelc, J. S., Langrock, C., Zhang, Q. & Fejer, M. M. Influence of domain disorder on parametric noise in quasi-phase-matched quantum frequency converters. *Opt. Lett.* **35**, 2804–2806 (2010).
- Bayer, M. *et al.* Fine structure of neutral and charged excitons in self-assembled In(Ga)As/(Al)GaAs quantum dots. *Phys. Rev. B* **65**, 195315 (2002).
- Xu, X. *et al.* Fast spin state initialization in a singly charged InAs-GaAs quantum dot by optical cooling. *Phys. Rev. Lett.* **99**, 097401 (2007).
- Xu, X. *et al.* Optically controlled locking of the nuclear field via coherent dark-state spectroscopy. *Nature* **459**, 1105–1109 (2009).
- Scully, M. O. & Drühl, K. Quantum eraser: a proposed photon correlation experiment concerning observation and “delayed choice” in quantum mechanics. *Phys. Rev. A* **25**, 2208–2213 (1982).
- Tanner, M. G. *et al.* Enhanced telecom wavelength single-photon detection with NbTiN superconducting nanowires on oxidized silicon. *Appl. Phys. Lett.* **96**, 221109 (2010).
- Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information* (Cambridge Univ. Press, 2000).
- Gao, W. B., Fallahi, P., Togan, E., Miguel-Sanchez, J. & Imamoglu, A. Observation of entanglement between a quantum dot spin and a single photon. *Nature* <http://dx.doi.org/10.1038/nature11573> (this issue).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank D. Press, T. Ladd, D. Sleiter, S. Tawfeeq, S. Rumley, D. Werthimer, A. Langman, C. Langrock, Q. Zhang, N. Namekata, S. Inoue, T. Inagaki and H. Kosaka for discussions, comments and technical assistance. We thank V. Zwiller and S. Dorenbos (TU Delft) for providing the superconducting detector samples used. This work was supported by the JSPS through its FIRST programme, NICT, NSF CCR-08 29694, NIST 60NANB9D9170, Special Coordination Funds for Promoting Science and Technology, and the State of Bavaria. J.S.P. and M.M.F. were supported by the United States AFOSR (grant FA9550-12-1-0110). Other authors were supported as follows: K.D.G. by a Herb and Jane Dwight Stanford Graduate Fellowship; P.L.M. by a David Cheriton Stanford Graduate Fellowship; J.S.P. by a Robert N. Noyce Stanford Graduate Fellowship; C.M.N. by a SU2P Entrepreneurial Fellowship; and R.H.H. by a Royal Society University Research Fellowship.

Author Contributions S.M., C.S., M.K. and S.H. grew and fabricated the samples. K.D.G. and Y.Y. designed the experiment. K.D.G., J.S.P., L.Y., P.L.M., C.M.N. and N.Y.K. performed the optical experiments. J.S.P. designed and fabricated the PPLN waveguides. J.S.P. and L.Y. developed the 2.2- μm set-up and the 1,560-nm filtering design. C.M.N. and R.H.H. packaged, characterized and implemented the SNSPD detectors. Y.Y., M.M.F., E.A. and A.F. guided the work. K.D.G. wrote the manuscript with input from all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.D.G. (kdegreve@stanford.edu).

METHODS

Device design. The quantum-dot sample used is similar to the ones used previously to investigate ultrafast optical control and single spin echoes⁷. Compared to those experiments, the dot density was reduced, and the wavelength used was blue-shifted (910 nm). An asymmetric, low- Q cavity (10 nm FWHM, centred at 910 nm) redirects the spontaneous emission preferentially in one direction, increasing collection efficiency and reducing required peak powers for coherent control⁷. An external magnetic field (Voigt geometry, perpendicular to the optical axis) splits both the electron spin states and the excited, trion states. The resulting Λ -systems are indicated in Fig. 1a, and the polarization selection rules are verified by means of polarization-selective photoluminescence (half-wave plate and polarizing beamsplitter, Fig. 1b).

Spin control and single-photon collection. The quantum-dot device is cooled to 1.6 K inside a superconducting magnetic cryostat (Oxford Spectromag; the magnetic field used varies between $B = 3$ and 6 T). A 0.68 NA aspheric lens inside the cryostat focuses the pump and rotation lasers onto the sample, which is scanned relative to the lens by means of slip-stick piezo-electric positioners (Attocube Systems). The coherent manipulation techniques used are identical to those reported previously⁷. A narrowband c.w. laser (New Focus Velocity) is used for spin initialization and readout, resonant to the $|\downarrow\rangle\text{--}|\uparrow\downarrow\uparrow\rangle$ -transition (910.10 nm for a 3 T magnetic field), and switched on and off by a fibre-based EOM (EOSpace). The interrogation time is 13 ns, with an initialization and readout fidelity of about 96% or higher (Supplementary Information). Coherent spin rotations are performed in a Ramsey-interferometry set-up, using pulses from a mode-locked laser (3-ps pulse duration, centre wavelength 911 nm, Spectra-Physics Tsunami), which are delayed relative to each other through a retroreflector on a motorized stage. Individual pulses are selected through free-space EOMs (Conoptics), which are double-passed in order to increase the extinction ratio. The selective excitation of the $|\uparrow\downarrow\uparrow\rangle$ -state before spontaneous emission is realized through a combination of optical pumping into the $|\uparrow\rangle$ -state, rotation by an optical π -pulse into the $|\downarrow\rangle$ -state, followed by the application of a 100-ps pulse (optical π -pulse) from another mode-locked laser (Spectra-Physics Tsunami), resonant with the $|\downarrow\rangle\text{--}|\uparrow\downarrow\uparrow\rangle$ -transition. This laser is synchronized with the rotation laser (Spectra-Physics Lok-to-Clock system). Accurate polarization control limits the probability of excitation into the $|\uparrow\downarrow\downarrow\rangle$ -state to less than 1%. Another fibre-optic modulator allows for pulse-picking of the 100-ps excitation pulses. The single-photon photoluminescence is collected in a confocal set-up, and split into two branches by a non-polarizing beamsplitter. One branch is cross-polarized with respect to the initialization and optical pumping lasers, and sent through a double-monochromator onto a single-photon counter for spin-state analysis (Perkin-Elmer SPCM; 20% quantum efficiency, 170 Hz ungated dark count rate). The other branch is sent to a polarization analysing stage (quarter- and half-wave plate and polarizer), after which it is coupled into single-mode fibres and sent to the downconversion set-up. The polarization analysing stage is carefully calibrated, in order to account for residual birefringence in the setup. All EOMs are controlled by mutually synchronized pulse-pattern generators (76 MHz Tektronix and 10 GHz Anritsu PPG), that are themselves synchronized to the repetition rate of the mode-locked lasers. In contrast to previous experiments, no software-based lock-in technique was used. Instead, spatial, polarization, wavelength and time-filtering are used to separate reflected light from the single photons.

Downconversion set-up. The 2.2- μm light pulses needed for conversion are generated by a difference-frequency generation (DFG) process that mixes the 3-ps, 911-nm pulses from the mode-locked laser with narrowband, c.w. 1,560-nm light in an MgO-doped, PPLN chip. The 1,560-nm light is modulated by a fibre-optic

modulator, and amplified by erbium-doped fibre amplifiers. After wave mixing, the residual 1,560-nm and 911-nm light is filtered out through a combination of dichroic and absorptive filters. The resulting pulse width depends on the exact power and wavelength used for the DFG process, but is measured to be between 3 and 8 ps.

A PPLN waveguide efficiently converts 910-nm, spontaneously emitted photons to 1,560-nm photons, conditional on overlap with the 2.2- μm pulses. Residual scattered light from the 910-nm and 2.2- μm branches is eliminated through a fibre-Bragg grating and a long-pass filter. The 1,560-nm photons are subsequently detected on an SNSPD, maintained at 2 K, with 14% system detection efficiency, 40-Hz ungated dark count rate and 100-ps FWHM timing jitter. Timing analysis is performed on a timing analyser (PicoQuant HydraHarp), used in time-tagged time-resolved (TTTR) mode, which allows for accurate gating of the signals of both the SPCM and the SNSPD in post-processing, thereby drastically reducing the effects of dark counts. The overall signal-to-noise ratio is limited by residual leakage of the 2.2- μm light, dark counts within the timing response of the SNSPD, and residual reflected light from the 100-ps excitation pulses, but is in general well below the single-photon level (signal-to-noise ratio ranging between 4:1 and 10:1 for the experiments described in this work). The timing of the 2.2- μm light is chosen such that the subsequent $\pi/2$ -pulse arrives well within the T_2^* -dephasing time of the quantum dot; similar results can be obtained for times up to microseconds by implementing a spin echo technique to overcome shot-to-shot dephasing⁷.

Data analysis and entanglement fidelity. The correlation data obtained in this work are the result of a histogram analysis, performed on the coincidence count rate between the downconverted single photons, and the single photons used for spin detection (Supplementary Information). The coincidence count rate is obtained through post-processing of the TTTR data stream, and comparing the coincidences within the same experimental run to those in subsequent, uncorrelated ones. The repetition rate is set at 39 or 52 ns, and the 0.1% single photon efficiency and time-gated frequency downconversion result in a 1,560-nm single-photon detection and entanglement generation rate of some 2–5 Hz. In combination with another 0.1% single-photon detection efficiency in order to detect the spin state, this results in an average coincidence rate of some 2–5 mHz. We emphasize that these losses are predominantly due to the inefficiency of extracting a single photon from the quantum dot, which can be significantly improved by accurate cavity design⁸. The conversion process in itself, while lossy due to the aggressive time-filtering in order to obtain good timing resolution, is rather effective, with internal quantum efficiencies estimated above 80%, and filtering losses of several dB maximum.

The entanglement fidelity analysis follows the same procedure used in the ion-trap⁴ and NV-diamond³ spin-photon entanglement experiments:

$$F \geq F_1 + F_2,$$

$$F_1 = \frac{1}{2} \left(\rho_{H\uparrow, H\uparrow} + \rho_{V\downarrow, V\downarrow} - 2\sqrt{\rho_{H\downarrow, H\downarrow}\rho_{V\uparrow, V\uparrow}} \right),$$

$$F_2 = \frac{1}{2} \left(\rho_{\sigma+\leftarrow, \sigma+\leftarrow} - \rho_{\sigma+\rightarrow, \sigma+\rightarrow} + \rho_{\sigma-\rightarrow, \sigma-\rightarrow} - \rho_{\sigma-\leftarrow, \sigma-\leftarrow} \right).$$

Here $\rho_{H\uparrow, H\uparrow}$ etc. refer to elements of the spin photon density matrix, which can be associated with the observed correlations in our experiment. For the values obtained in our work, $F_1 = 0.44 \pm 0.06$, $F_2 = 0.36 \pm 0.06$, from which we obtain a bound on the quoted fidelity: $F \geq 0.8 \pm 0.085$.

Observation of entanglement between a quantum dot spin and a single photon

W. B. Gao¹, P. Fallahi¹, E. Togan¹, J. Miguel-Sanchez¹ & A. Imamoglu¹

Entanglement has a central role in fundamental tests of quantum mechanics¹ as well as in the burgeoning field of quantum information processing. Particularly in the context of quantum networks and communication, a main challenge is the efficient generation of entanglement between stationary (spin) and propagating (photon) quantum bits². Here we report the observation of quantum entanglement between a semiconductor quantum dot spin and the colour of a propagating optical photon. The demonstration of entanglement relies on the use of fast, single-photon detection, which allows us to project the photon into a superposition of red and blue frequency components. Our results extend the previous demonstrations of single-spin/single-photon entanglement in trapped ions³, neutral atoms^{4,5} and nitrogen-vacancy centres⁶ to the domain of artificial atoms in semiconductor nanostructures that allow for on-chip integration of electronic and photonic elements^{7,8}. As a result of its fast optical transitions and favourable selection rules, the scheme we implement could in principle generate nearly deterministic entangled spin-photon pairs at a rate determined ultimately by the high spontaneous emission rate. Our observation constitutes a first step towards implementation of a quantum network with nodes⁹ consisting of semiconductor spin quantum bits^{10–12}.

Recent advances have demonstrated the potential of semiconductor spin quantum bits (qubits) in the implementation of fast, single-qubit^{13–16} and two-qubit¹⁷ quantum gates. Realization of a large-scale, spin-based quantum information processor, however, remains an outstanding challenge owing to limitations on the coherent coupling of distant qubits. An alternative route is provided by the concept of distributed quantum computation, which involves small-scale, few-qubit quantum processor nodes interconnected by single photons that are quantum correlated with spin qubits¹⁸. Remarkably, photonic circuits allowing for efficient optical-photon-based quantum communication between the nodes can be fabricated on the same semiconductor chip⁷. For this distributed approach to solid-state quantum information processing, as well as for long-distance quantum communication, generation of the spin-photon entanglement that we demonstrate here emerges as a key enabling technology.

Our experiments utilize a single-electron-charged, self-assembled InGaAs quantum dot. Figure 1a shows the relevant energy-level diagram as well as the allowed optical transitions in Voigt geometry, where an external magnetic field (B_x) is applied perpendicular to the growth direction¹⁹. In our scheme, it is possible to generate and verify high-fidelity entanglement for $0.2\text{ T} \leq B_x \leq 1\text{ T}$: the lower limit ensures that the quantum dot is optically excited only to the red trion state, $|T_r\rangle$ (Fig. 1a), and the upper limit is dictated by the detector time resolution. In our experiments, we choose $B_x = 0.7\text{ T}$. All four transitions between the two ground states and the two excited states (Fig. 1a) have identical oscillator strengths and are linearly polarized either parallel (H) or perpendicular (V) to the applied field direction¹⁹. The ground states of the quantum dot are identified by the orientation of the electron spin, with $|\uparrow\rangle$ and $|\downarrow\rangle$ respectively denoting spins parallel and antiparallel to the magnetic field direction. Spontaneous emission

of an H- or V-polarized photon at frequency ω_{red} or, respectively, ω_{blue} from the trion state $|T_r\rangle$ at rate $\Gamma/2$ brings the quantum dot back into the state $|\downarrow\rangle$ or, respectively, $|\uparrow\rangle$. In addition to resonant laser excitation of the trion states, our scheme uses non-resonant, right-hand-circularly (σ^+)-polarized, 4-ps-long mode-locked laser pulses that are red-detuned from the trion transitions by $\sim 210\text{ GHz}$: these pulses induce a coherent rotation of the electron spin, with minimal trion excitation¹³. We depict these pulses in Fig. 1a with a time-dependent effective magnetic field along the z direction (Supplementary Information).

The basic principle behind the deterministic generation of a spin-photon entangled state is straightforward: following the excitation of the quantum dot into the trion state, the radiative recombination projects the coupled quantum dot/photonic mode system into the entangled state

$$|\Psi\rangle = \frac{1}{\sqrt{2}}(|\downarrow\rangle|\omega_{\text{red}}; \text{H}\rangle + i|\uparrow\rangle|\omega_{\text{blue}}; \text{V}\rangle) \quad (1)$$

written in the computational basis, where $|\omega_{\text{red}}; \text{H}\rangle$ and $|\omega_{\text{blue}}; \text{V}\rangle$ denote propagating single-photon-pulse state vectors with respective centre frequencies ω_{red} and ω_{blue} and respective polarizations H and V. The electronic Zeeman energy satisfies $\omega_Z = \omega_{\text{blue}} - \omega_{\text{red}}$.

We use resonance fluorescence measurements^{20,21} to map the optical transitions of the quantum dot. Observation of resonance fluorescence from the quantum dot transitions in Voigt geometry requires the presence of a second laser field that rotates the electron spin, because otherwise spin pumping, which takes place within a few nanoseconds¹⁹, ensures that the quantum dot remains dark to optical excitation. To characterize the quantum dot, we use a pulse sequence where two resonant laser pulses, respectively 5 and 1.2 ns long, are applied in addition to a 4-ps pulse implementing spin rotation (Fig. 1b). The resonance fluorescence spectrum that we obtain as we scan the resonant laser across the four allowed quantum dot transitions is depicted in Fig. 1c. Our fast, single-photon detection capability allows us to monitor the time dependence of the resonance fluorescence: as Fig. 1d shows, we can easily distinguish photons that are emitted during excitation by the resonant laser both from those emitted as a result of (unwanted) trion excitation induced by the π -pulse and the background photons from the 4-ps laser. This ability to monitor photons generated in definite time windows is crucial for the verification of spin-photon entanglement. The experimental set-up that allows us to implement two-laser excitation as well as selective detection of resonance fluorescence generated in different time windows or at different frequencies, or both, is depicted in Fig. 1e (Methods Summary). Suppression of the σ^+ -polarized background laser light is ensured by a polarizer²¹ that projects the polarization of the transmitted resonance fluorescence photons to σ^- (left-hand circular). Because the spin polarization correlations are erased at this stage, entanglement is between the electron spin orientation and the centre frequency of the single-photon pulse (such that the photonic states can be written as $|\omega_{\text{red}}\rangle$ and $|\omega_{\text{blue}}\rangle$).

To demonstrate spin-photon entanglement, we first prepare the quantum dot in state $|\uparrow\rangle$ with a probability exceeding 87% by applying a 5-ns resonant laser field tuned to the $|\downarrow\rangle \leftrightarrow |T_r\rangle$ transition; we refer to

¹Institute of Quantum Electronics, ETH Zurich, CH-8093 Zurich, Switzerland.

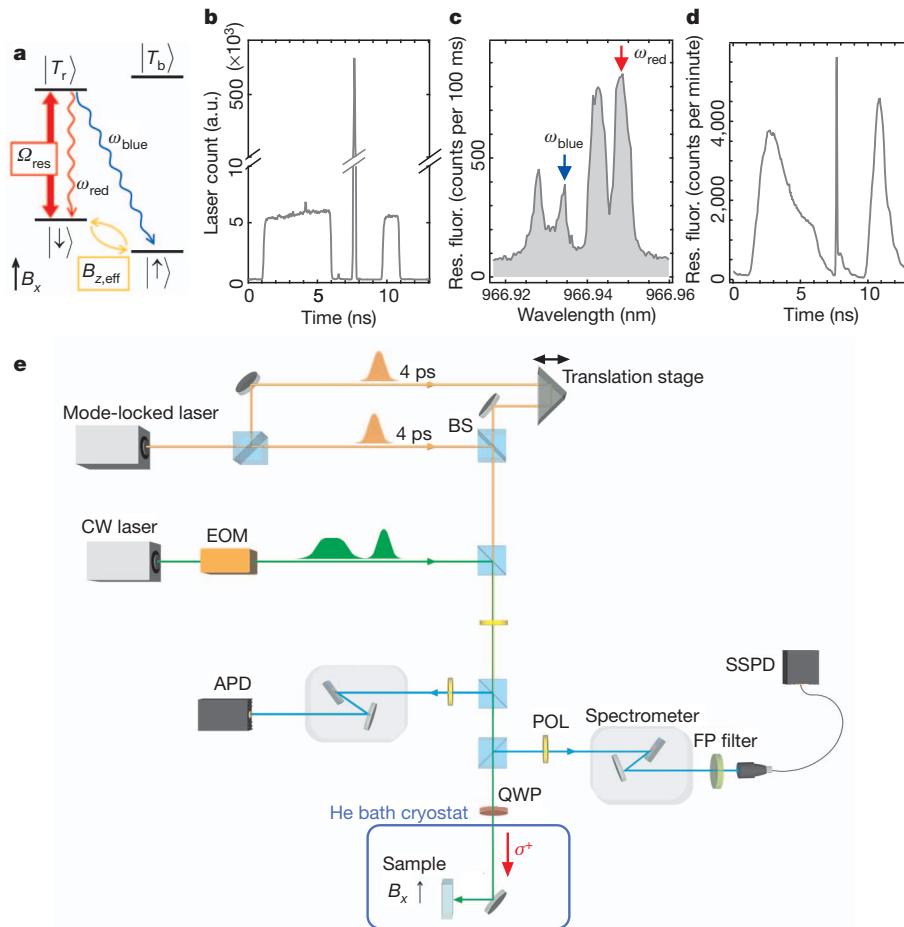


Figure 1 | Quantum dot transitions and experimental set-up. **a**, The energy-level diagram of a single-electron-charged InGaAs quantum dot in Voigt geometry. Spontaneous emission following resonant optical excitation of the trion state, $|T_r\rangle$, with Rabi frequency Ω_{res} leads to the generation of an entangled spin–photon state. Non-resonant, σ^+ -polarized, 4-ps laser pulses implement coherent spin rotation; the effect of such a laser pulse is equivalent to a time-dependent magnetic field, $B_{z,\text{eff}}$, that is perpendicular to the external field B_x (Supplementary Information). **b**, The pulse sequence used to characterize the quantum dot transitions: following a 5-ns resonant laser excitation, a 4-ps π -pulse is applied. A 1.2-ns resonant laser excitation

completes the pulse sequence, which has a repetition period of 13 ns. **c**, Resonance fluorescence spectrum generated by the pulse sequence depicted in **b** as the laser frequency ω_L is scanned across the quantum dot transitions at $B_x = 0.7$ T. **d**, Time dependence of resonance fluorescence obtained when the laser is on resonance with the $|\downarrow\rangle \leftrightarrow |T_r\rangle$ transition ($\omega_L = \omega_{\text{red}}$), indicated by the red arrow in **c**. The spike that coincides with the 4-ps laser pulse is due to incomplete suppression of reflected laser photons. **e**, The schematic of the experimental set-up highlighting the important optical elements. BS, beam splitter; CW, continuous wave; EOM, electro-optic modulator; FP, Fabry–Pérot; POL, polarizer; QWP, quarter-wave plate.

this laser as the measurement/preparation pulse. We then transfer the quantum dot to the $|\downarrow\rangle$ state using a 4-ps π -pulse¹³. An entangled spin–photon pair is then generated by exciting $|T_r\rangle$ with a 1.2-ns resonant laser pulse, which we refer to as the entanglement pulse; the corresponding pulse sequence is depicted in Fig. 2a. The whole pulse sequence is repeated after 13 ns, with the 5-ns resonant laser implementing spin measurement for the preceding pulse sequence as well as ensuring spin pumping/preparation in the $|\uparrow\rangle$ state for the subsequent cycle (Fig. 2a). The fact that the average number of photons emitted during this spin measurement/pumping time window is ~ 1 , together with our relatively modest combined collection and detection efficiency of $\sim 0.1\%$, implies that we extract heralded spin information²²; that is, the unlikely event of a photon being detected at the avalanche photodiode (APD) tells us with very high confidence level that the spin (before the measurement) is $|\downarrow\rangle$, whereas the absence of a detection event yields no information. An additional π -pulse (Fig. 2a, dashed curve) applied 3 ns after the onset of the entanglement pulse enables a photon detection event at the APD during the measurement/preparation pulse to project the spin into $|\uparrow\rangle$ rather than $|\downarrow\rangle$.

Figure 2b shows the time-resolved resonance fluorescence at frequencies ω_{blue} and ω_{red} . The depicted resonance fluorescence starts at

the onset of the entanglement pulse, which we take to be $t = 0$ in Figs 2 and 3. For all experiments depicted in Fig. 2, we use a Fabry–Pérot interferometer to select out photons either at ω_{blue} or ω_{red} before sending them to the superconducting single-photon detector (SSPD). The (unconditional) resonance fluorescence counts at the two frequencies are nearly identical in intensity, indicating that the two decay channels have equal oscillator strengths. The shape of these single-photon pulses is determined by the time dependence of the trion population, which in turn is determined by the excitation laser (through its pulse shape and the Rabi frequency, Ω_{res}) and the spontaneous emission time $\Gamma^{-1} = 650$ ps. Figure 2c (green bars) shows the normalized, zero-time-delay photon autocorrelation function, $g^{(2)}(0) = 0.02 \pm 0.01$, for photons emitted during the time interval [1.2 ns, 1.64 ns] (Fig. 2b, dashed green box). This measurement demonstrates that the unlikely two-photon emission events lead to a small degradation of the entanglement. When we measure photon correlations for the whole single-photon pulse, we find that $g^{(2)}(0) = 0.16 \pm 0.01$ (Fig. 2c, black bars). This value is limited partly by a residual contribution from the reflected laser photons and partly by multiple excitation of the quantum dot during the entanglement pulse. We note that verifying the spin–photon entanglement in the time interval [1.2 ns, 1.64 ns] also allows us to

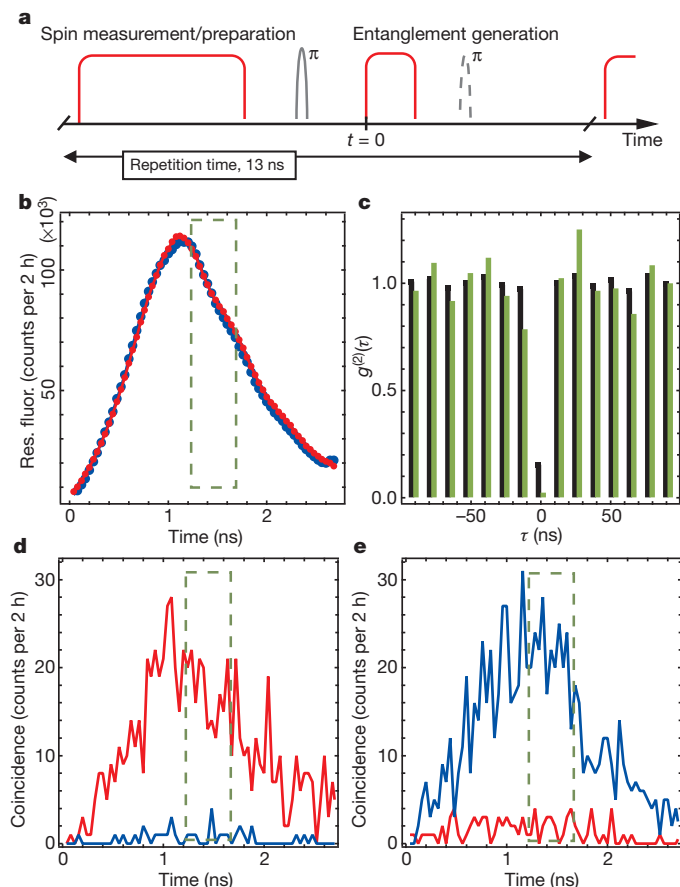


Figure 2 | Measurement of classical spin-photon correlations. **a**, The pulse sequence used to measure the spin-photon correlations in the computational (non-rotated) basis. A 5-ns laser excitation resonant with the $|\downarrow\rangle \leftrightarrow |T_r\rangle$ transition implements spin pumping/preparation in $|\uparrow\rangle$; detection of a resonance fluorescence event during this measurement/preparation pulse projects the spin state to $|\downarrow\rangle$. Following a 4-ps-long π -pulse, a 1.2-ns laser, resonant with the $|\downarrow\rangle \leftrightarrow |T_r\rangle$ transition, initiates the generation of an entangled spin-photon pair. An additional π -pulse (dashed curve) is applied only to obtain the data shown in **e**. **b**, Time-resolved resonance fluorescence obtained at ω_{red} (red) and ω_{blue} (blue) using the pulse sequence shown in **a**: identical intensities and pulse shapes at the two frequencies indicate that the detected events stem from a single trion state ($|T_r\rangle$) and that the oscillator strength of the transitions at ω_{red} and ω_{blue} are identical. **c**, Photon correlation measurement carried out on resonance fluorescence generated in a 3-ns time window starting at the onset of the entanglement pulse ($t = 0$). The observed $g^{(2)}(0) = 0.16$ (black) is limited by background laser photons. When we carry out photon correlation measurements on photons detected during the time interval [1.2 ns, 1.64 ns] (box in **b**), we find that $g^{(2)}(0) = 0.02$ (green). **d**, Time-dependent resonance fluorescence at ω_{red} (red) and ω_{blue} (blue), conditioned on detection of a photon during the subsequent measurement/preparation pulse. The strong suppression of conditional photon detection events at ω_{blue} demonstrates the strong classical correlation between spin measurement yielding $|\downarrow\rangle$ and a photon detection at ω_{red} . **e**, Same as in **d** but now with a π -pulse applied before the measurement/preparation pulse, showing strong correlations between spin detection in $|\uparrow\rangle$ and photon detection at ω_{blue} .

avoid the effects of the fluctuating nuclear spin environment on the entanglement fidelity.

Demonstration of spin-photon correlations in the computational/non-rotated basis requires the measurement of classical correlation between the spin direction and the photon frequency. This is achieved by carrying out a coincidence measurement between resonance fluorescence photons at either ω_{blue} or ω_{red} generated immediately after the entanglement pulse and those generated during the measurement/preparation pulse. Conditional on the detection of the spin state $|\downarrow\rangle$, we find that the probability of detecting a blue (ω_{blue}) photon at the SSPD

is drastically suppressed compared with the probability of detecting a red (ω_{red}) photon (Fig. 2d). The ratio of the red to blue photons in the time interval [1.2 ns, 1.64 ns] is $\sim 21:3$. To quantify completely the classical correlations in the computational basis, we introduce a 4-ps π -pulse at $t = 3$ ns that makes it possible to condition the single-photon colour measurements on detection of the spin in $|\uparrow\rangle$. In this case, we find that the red-photon detection events are strongly suppressed (Fig. 2e). The ratio of the blue to red photons in this case, however, is only $\sim 10:3$, owing to (unwanted) trion excitation by the 4-ps laser that leads to lower fidelity in the subsequent spin measurement. From these two measurements, we determine the fidelity of the classical correlations to be $F_1 = 0.87 \pm 0.09$ (Supplementary Information).

To verify that the generated spin-photon state is entangled, we next show that the correlations between spin orientation and photon colour exist in a rotated basis⁶. In the rotated basis, the relative phase between the two components of the entangled state in equation (1) is crucial. After generation at time t_g , the two components acquire a time-dependent phase that stems from the different propagation phase factors of the blue and red frequency components of the photonic mode. In our experiments, we infer the generation time from the detection time, t_d , of the single photon at the SSPD, through the relation $t_g = t_d - L/c$, where L denotes the distance from the quantum dot to the detector and c is the speed of light. Hence, given a photon detection event at t_d , we can a-posteriori write the time dependence of the spin-photon entangled state of equation (1) as

$$|\Psi(t)\rangle = \frac{1}{\sqrt{2}} (|\downarrow\rangle|\omega_{\text{red}}; H\rangle e^{-i\omega_Z(t-t_g)} + i|\uparrow\rangle|\omega_{\text{blue}}; V\rangle)$$

The requisite spin measurement is implemented by rotating the quantum dot electron spin with either a $\pi/2$ -pulse or a $3\pi/2$ -pulse at $t = t_1 = 1.7$ ns. A photon detection event at the APD during the following measurement/preparation pulse then projects the electron spin into $(|\downarrow\rangle - i|\uparrow\rangle)/\sqrt{2}$ ($\pi/2$ -pulse) or $(|\downarrow\rangle + i|\uparrow\rangle)/\sqrt{2}$ ($3\pi/2$ -pulse). The photonic wavefunction consistent with the spin measurement after the $\pi/2$ -pulse can be written as $(|\omega_{\text{red}}; H\rangle e^{-i\omega_Z(t_1-t_g)} - |\omega_{\text{blue}}; V\rangle)/\sqrt{2}$ (Supplementary Information). After passing through the polarizer, which fixes the polarization of the photon to be $(|H\rangle - i|V\rangle)/\sqrt{2}$, this wavefunction is

$$|\tilde{\Phi}\rangle = \frac{1}{\sqrt{2}} (|\omega_{\text{red}}\rangle e^{-i\omega_Z(t_1-t_g)} - i|\omega_{\text{blue}}\rangle) \quad (2)$$

In the experiment, we remove the Fabry-Pérot interferometer and directly send the single-photon pulses onto a fast SSPD²³ with a jitter of $\Delta\tau = 60$ ps. With a SSPD that had jitter smaller than $1/\omega_Z$, we could determine t_g well enough to resolve the time dependence in equation (2), which in turn would indicate that the photonic state is projected into a superposition of its two frequency components. The resulting coincidence probability is $P \approx [1 + \sin(\omega_Z(t_1 - t_g))]/2$. As a consequence of the time-dependent relative phase, two single-photon states generated at different times, t_g^a and t_g^b , such that $\omega_Z(t_g^a - t_g^b) = \pi$, allow us to infer the overlap of the photonic state with two orthogonal states. Random photon detection times together with the fact that the single-photon pulse width, τ , satisfies $\omega_Z\tau > 1$, allow us effectively to implement projective measurements into orthogonal states such as $|\omega_{\text{red}}\rangle \pm |\omega_{\text{blue}}\rangle$.

Figure 3b shows the coincidence between the single-photon detection events induced by the entanglement pulse and the detection of a photon during the measurement/preparation pulse, following a $\pi/2$ -pulse: the oscillations with a period given by $2\pi/\omega_Z = 238$ ps in the conditioned single-photon detection events (Fig. 3b, red squares) are in stark contrast to the unconditioned single-photon detection events shown in Fig. 2b. This particular time dependence, stemming from the conditioned photonic wavefunction given in equation (2), constitutes a remarkable manifestation of the quantum coherence of

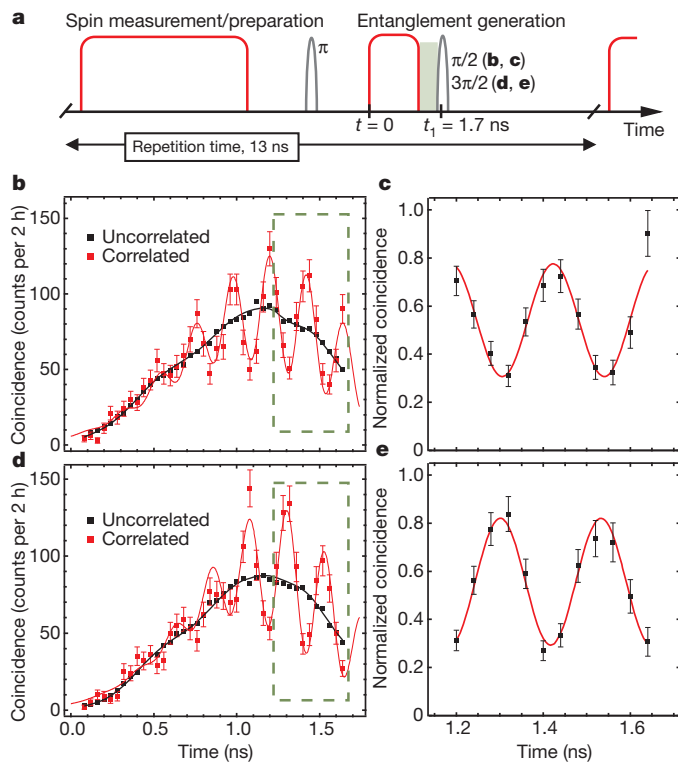


Figure 3 | Demonstration of quantum correlation between the electron spin and the emitted single-photon pulse. **a**, The pulse sequence used to measure the spin–photon correlations in the rotated basis: the only difference from Fig. 2a is the introduction of a $\pi/2$ - or $3\pi/2$ -pulse, applied at $t = 1.7$ ns, that rotates the spin measurement basis such that the detection of a photon during the subsequent measurement/preparation pulse projects the spin to $|\downarrow\rangle - i|\uparrow\rangle$ or, respectively, $|\downarrow\rangle + i|\uparrow\rangle$. The shading indicates the [1.2 ns, 1.64 ns] entanglement analysis time window. **b**, Time-resolved coincidence events between the single-photon detection following the entanglement pulse and the detection of a photon during the first measurement/preparation pulse following a $\pi/2$ -pulse (red squares). For comparison, an average of coincidence events between the spin and photon detection events taking place in 25 different excitation/preparation cycles is also shown (black squares). The red curve is a fit to the experimental data, and the black curve is the rescaled single-photon pulse shown in Fig. 2b. **c**, Normalized coincidence during the [1.2 ns, 1.64 ns] window (box in **b**); normalization is carried out by dividing the number of counts from correlated spin–photon pairs by twice the number of uncorrelated events in the corresponding time bin. **d**, Same as in **b** but now with a $3\pi/2$ -pulse. **e**, Same as in **c** but now with a $3\pi/2$ -pulse. All error bars represent 1 s.d., deduced from Poissonian statistics of the raw detection events.

the entangled spin–photon system. For comparison, we also show an average of coincidence events between the spin and photon detection events taking place in different excitation/preparation cycles (Fig. 3b, black squares). As expected, these coincidences do not show any oscillations, but they help us quantify the degree of quantum correlation. Normalizing the coincidence detection between correlated spin–photon pairs and focusing on the time interval between 1.2 and 1.64 ns, we find a visibility of about $45 \pm 6\%$ for the observed oscillations (Fig. 3c). Figure 3d shows the time-dependent coincidence measurements obtained by applying a $3\pi/2$ -pulse before the measurement/preparation pulse. Figure 3e in turn shows that when we focus on the same time interval, the oscillations for spin detection along $(|\downarrow\rangle + i|\uparrow\rangle)/\sqrt{2}$ (Fig. 3e) are π out of phase relative to those along $(|\downarrow\rangle - i|\uparrow\rangle)/\sqrt{2}$ (Fig. 3c) and have a visibility of $47 \pm 4\%$. The actual coincidence detection rate in the detection time window [1.2 ns, 1.64 ns] is 240 counts per hour. Using the data from Fig. 3c, e, we determine the fidelity of the quantum correlations in this rotated basis to be the average of these two visibilities, yielding $F_2 = 0.46 \pm 0.04$ (Supplementary Information). The overall measured entanglement fidelity is then

$F \geq (F_1 + F_2)/2 = 0.67 \pm 0.05$, limited predominantly by the finite jitter of the SSPD.

A major limitation in the demonstration of quantum correlations between the electron spin and the photon frequency stems from the relatively short ground-state spin decoherence time, T_2^* , of the electron spin, originating from the slowly fluctuating nuclear spin environment. The decay of the oscillations as a function of $t_1 - t_g$ in conditioned single-photon detection events (Fig. 3b, d) constitutes a measurement of the electron spin decoherence that was previously carried out using Ramsey interferometry²⁴. By fitting the decay of coincidence measurements in the time window [0 ns, 1.64 ns] we find that $T_2^* = 1.1 \pm 0.2$ ns for the data in Fig. 3b and that $T_2^* = 0.9 \pm 0.2$ ns for the data in Fig. 3d, for this quantum dot.

In our experiments, the extraction efficiency of the spin–photon entanglement (that is, the fraction of emitted photon collected by the objective) is ~ 0.1 . We emphasize, however, that near-deterministic generation of entanglement is possible using InGaAs quantum dots: several groups have demonstrated that quantum dot emissive properties can be drastically improved by incorporating them in tailored photonic nanostructures^{7,25,26}. The fraction of photons emitted at the zero-phonon line typically exceeds 90% in self-assembled quantum dots; enhanced radiative recombination rate in devices with Purcell factors much greater than one ensures that the fraction of photons emitted at the zero-phonon line approaches unity⁸. By embedding a quantum dot in a micropillar cavity yielding a Purcell factor exceeding one, it is also possible to ensure single-photon collection efficiency approaching unity²⁶.

Our results realize the long-standing goal of establishing a light–matter quantum interface in optically active quantum dots in GaAs-based materials^{2,11}. Integration of self-assembled or other optically active quantum dots with spin qubits in electrically gated quantum dots¹² would allow entangled spin–photon pairs to be used to entangle distant spin qubits. The relatively short T_2^* observed in our experiment presents a limitation on the separation of distant spin qubits that can be entangled. We emphasize that methods for narrowing the fluctuating Overhauser field²⁷ could be used to increase T_2^* . Implementing such a scheme before spin–photon entanglement generation or using spin-echo techniques would extend the distance limit set by T_2^* . Observation of indistinguishable photons from two distant, self-assembled quantum dots in a Hong–Ou–Mandel interferometer, which was demonstrated experimentally^{28,29}, constitutes an additional key step towards realization of distant-spin entanglement. We emphasize that coincidence detection in a Hong–Ou–Mandel set-up that signals successful entanglement swapping can be implemented in a straightforward way using photons from two spin–frequency entangled pairs³⁰.

METHODS SUMMARY

Experimental set-up. Two time-delayed 4-ps pulses from a mode-locked laser are combined with a 5-ns pulse and a 1.2-ns pulse obtained from a continuous-wave laser using an electro-optic modulator, and the resulting single beam is sent to the quantum dot. The relative delay of the 4-ps pulses can be changed with a translation stage. A polarizer and a quarter-wave plate ensure that the lasers are σ^+ -polarized at the quantum dot. Emitted quantum dot photons are detected by an APD and a SSPD through two collection paths. In each path, a combination of a polarizer and a spectrometer suppresses the reflected laser background. A Fabry–Pérot filter is used to select the colour of the entanglement photon for the computational basis measurements and is removed for the rotated basis measurements. **Synchronization.** In the experiment, the fixed repetition rate of a mode-locked laser is used as a reference to synchronize the time delay of experimental pulses and also to synchronize the time of detection events. The repetition rate is measured using a fast photodiode (40-ps rise time). The 32nd harmonic of the repetition rate is extracted with a band-pass filter centred at 2.4 GHz, amplified (70 dB gain) and then fed into the clock input of a pulse-pattern generator (PPG), which is used to synchronize both the pulse generation and the detection circuitries. An electrical pulse output from the PPG is used to drive an electro-optic modulator, modulating a single-mode continuous-wave laser and ensuring that the spin measurement/preparation pulses and entanglement-generation pulses are always in a fixed time order. Another output of the PPG is used to pre-scale the 2.4-GHz

synchronization signal, before being fed into the clock input of a time-correlated single-photon-counting module. This module records the time delay of photon arrival events (via the APD or the SSPD) relative to the pre-scaled clock.

Received 7 June; accepted 10 September 2012.

- Aspect, A., Grangier, P. & Roger, G. Experimental realization of Einstein-Podolsky-Rosen-Bohm gedanken experiment: a new violation of Bell's inequalities. *Phys. Rev. Lett.* **49**, 91–94 (1982).
- DiVincenzo, D. P. The physical implementation of quantum computation. *Fortschr. Phys.* **48**, 771–783 (2000).
- Blinov, B., Moehring, D. & Duan, L. Observation of entanglement between a single trapped atom and a single photon. *Nature* **428**, 153–157 (2004).
- Volz, J. *et al.* Observation of entanglement of a single photon with a trapped atom. *Phys. Rev. Lett.* **96**, 030404 (2006).
- Wilk, T., Webster, S. C., Kuhn, A. & Rempe, G. Single-atom single-photon quantum interface. *Science* **317**, 488–490 (2007).
- Togan, E. *et al.* Quantum entanglement between an optical photon and a solid-state spin qubit. *Nature* **466**, 730–734 (2010).
- Hennessy, K. *et al.* Quantum nature of a strongly coupled single quantum dot-cavity system. *Nature* **445**, 896–899 (2007).
- Faraon, A. *et al.* Coupling of nitrogen-vacancy centers to photonic crystal cavities in monocrystalline diamond. *Phys. Rev. Lett.* **109**, 033604 (2012).
- Yuan, Z.-S. *et al.* Experimental demonstration of a BDCZ quantum repeater node. *Nature* **454**, 1098–1101 (2008).
- Loss, D. & DiVincenzo, D. P. Quantum computation with quantum dots. *Phys. Rev. A* **57**, 120–126 (1998).
- Imamoğlu, A. *et al.* Quantum information processing using quantum dot spins and cavity QED. *Phys. Rev. Lett.* **83**, 4204–4207 (1999).
- Hanson, R. *et al.* Spins in few-electron quantum dots. *Rev. Mod. Phys.* **79**, 1217–1265 (2007).
- Press, D., Ladd, T. D., Zhang, B. & Yamamoto, Y. Complete quantum control of a single quantum dot spin using ultrafast optical pulses. *Nature* **456**, 218–221 (2008).
- Greilich, A. *et al.* Ultrafast optical rotations of electron spins in quantum dots. *Nature Phys.* **5**, 262–266 (2009).
- Poem, E. *et al.* Accessing the dark exciton with light. *Nature Phys.* **6**, 993–997 (2010).
- Kim, D. *et al.* Ultrafast optical control of entanglement between two quantum-dot spins. *Nature Phys.* **7**, 223–229 (2011).
- Shulman, M. D. *et al.* Demonstration of entanglement of electrostatically coupled singlet-triplet qubits. *Science* **336**, 202–205 (2012).
- Cirac, J., Ekert, A. & Huelga, S. Distributed quantum computation over noisy channels. *Phys. Rev. A* **59**, 4249–4254 (1999).
- Xu, X. *et al.* Fast spin state initialization in a singly charged InAs-GaAs quantum dot by optical cooling. *Phys. Rev. Lett.* **99**, 097401 (2007).
- Müller, A. *et al.* Resonance fluorescence from a coherently driven semiconductor quantum dot in a cavity. *Phys. Rev. Lett.* **99**, 187402 (2007).
- Vamivakas, A. N., Zhao, Y., Lu, C.-Y. & Atatüre, M. Spin-resolved quantum-dot resonance fluorescence. *Nature Phys.* **5**, 198–202 (2009); corrigendum **5**, 925 (2009).
- Yilmaz, S. T., Fallahi, P. & Imamoglu, A. Quantum-dot-spin single-photon interface. *Phys. Rev. Lett.* **105**, 033601 (2010).
- Gol'tsman, G. N. *et al.* Picosecond superconducting single-photon detector. *Appl. Phys. Lett.* **79**, 705–707 (2001).
- Press, D. *et al.* Ultrafast optical spin echo in a single quantum dot. *Nature Photon.* **4**, 367–370 (2010).
- Claudon, J. *et al.* A highly efficient single-photon source based on a quantum dot in a photonic nanowire. *Nature Photon.* **4**, 174–177 (2010).
- Dousse, A. *et al.* Ultrabright source of entangled photon pairs. *Nature* **466**, 217–220 (2010).
- Xu, X. *et al.* Optically controlled locking of the nuclear field via coherent dark-state spectroscopy. *Nature* **459**, 1105–1109 (2009).
- Flagg, E. B. *et al.* Interference of single photons from two separate semiconductor quantum dots. *Phys. Rev. Lett.* **104**, 137401 (2010).
- Patel, R. *et al.* Two-photon interference of the emission from electrically tunable remote quantum dots. *Nature Photon.* **4**, 632–635 (2010).
- Moehring, D. L. *et al.* Entanglement of single-atom quantum bits at a distance. *Nature* **449**, 68–71 (2007).

Supplementary Information is available in the online version of the paper.

Acknowledgements We acknowledge discussions with W. Chin, M. Kroner, A. Srivastava, J. Elzerman, A. Reinhard, T. Volz, P. Maletinsky and D. Gershoni. This work is supported by NCCR Quantum Science and Technology, a research instrument of the Swiss National Science Foundation; the Swiss NSF (grant no. 200021-140818); an ERC Advanced Investigator Grant (A.I.); and a Marie Curie International Incoming Fellowship within FP7 (W.B.G.).

Author Contributions All authors contributed extensively to this work.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to W.B.G. (weibo@phys.ethz.ch) or A.I. (imamoglu@phys.ethz.ch).

Spontaneous motion in hierarchically assembled active matter

Tim Sanchez^{1*}, Daniel T. N. Chen^{1*}, Stephen J. DeCamp^{1*}, Michael Heymann^{1,2} & Zvonimir Dogic¹

With remarkable precision and reproducibility, cells orchestrate the cooperative action of thousands of nanometre-sized molecular motors to carry out mechanical tasks at much larger length scales, such as cell motility, division and replication¹. Besides their biological importance, such inherently non-equilibrium processes suggest approaches for developing biomimetic active materials from microscopic components that consume energy to generate continuous motion^{2–4}. Being actively driven, these materials are not constrained by the laws of equilibrium statistical mechanics and can thus exhibit sought-after properties such as autonomous motility, internally generated flows and self-organized beating^{5–7}. Here, starting from extensile microtubule bundles, we hierarchically assemble far-from-equilibrium analogues of conventional polymer gels, liquid crystals and emulsions. At high enough concentration, the microtubules form a percolating active network characterized by internally driven chaotic flows, hydrodynamic instabilities, enhanced transport and fluid mixing. When confined to emulsion droplets, three-dimensional networks spontaneously adsorb onto the droplet surfaces to produce highly active two-dimensional nematic liquid crystals whose streaming flows are controlled by internally generated fractures and self-healing, as well as unbinding and annihilation of oppositely charged disclination defects. The resulting active emulsions exhibit unexpected properties, such as autonomous motility, which are not observed in their passive analogues. Taken together, these observations exemplify how assemblages of animate microscopic objects exhibit collective biomimetic

properties that are very different from those found in materials assembled from inanimate building blocks, challenging us to develop a theoretical framework that would allow for a systematic engineering of their far-from-equilibrium material properties.

We assembled active materials from microtubule filaments, which are stabilized with the non-hydrolysable nucleotide analogue GMPCPP, leading to an average length of 1.5 μm . Bundles were formed by adding a non-adsorbing polymer—poly(ethylene glycol) or PEG—which induces attractive interactions through the well-studied depletion mechanism. To drive the system far from equilibrium, we added biotin-labelled fragments of kinesin-1, a molecular motor that converts chemical energy from ATP hydrolysis into mechanical movement along a microtubule⁸. Kinesins were assembled into multi-motor clusters by tetrameric streptavidin, which can simultaneously bind and move along multiple microtubules, inducing inter-filament sliding (Fig. 1a). In this respect, our experiments build upon important earlier work that demonstrated the formation of asters and vortices in networks of unbundled microtubules and kinesin^{9,10}. However, compared to these dispersed networks, the proximity and alignment of depletion-bundled microtubules greatly increases the probability of kinesin clusters simultaneously binding and walking along neighbouring filaments, thus enhancing the overall activity.

Motor-induced sliding of aligned microtubules depends on their relative polarity. Kinesin clusters generate sliding forces between microtubules of opposite polarity, whereas no sliding force is induced between microtubules of the same polarity^{11–13}. To study the dynamics

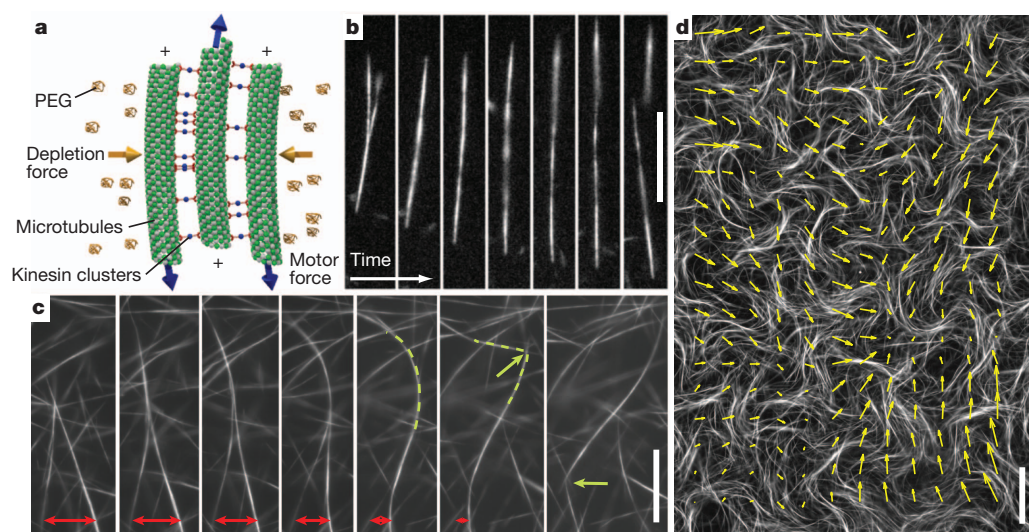


Figure 1 | Active microtubule networks exhibit internally generated flows. **a**, Schematic illustration of an extensile microtubule-kinesin bundle, the basic building block used for the assembly of active matter. Kinesin clusters exert inter-filament sliding forces, whereas depleting PEG polymers induce microtubule bundling. **b**, Two microtubule bundles merge and the resultant bundle immediately extends, eventually falling apart. Time interval, 5 s; scale bar, 15 μm . **c**, In a percolating microtubule network, bundles constantly merge (red arrows), extend, buckle (green dashed lines), fracture, and self-heal to produce a robust and highly dynamic steady state. Time interval, 11.5 s; scale bar, 15 μm . **d**, An active microtubule network viewed on a large scale. Arrows indicate local bundle velocity direction. Scale bar, 80 μm .

¹Martin Fisher School of Physics, Brandeis University, 415 South Street, Waltham, Massachusetts 02454, USA. ²Graduate Program in Biophysics and Structural Biology, Brandeis University, 415 South Street, Waltham, Massachusetts 02454, USA.

*These authors contributed equally to this work.

of active bundles, we confined a dilute suspension of microtubule bundles to quasi-two-dimensional (2D) chambers. No sliding is observed for isolated bundles, indicating that they are locally polarity sorted. However, owing to passive diffusion, quasi-static bundles occasionally encounter each other and join as a result of attractive depletion interactions. Merging bundle domains are equally likely to have the same or opposite polarity, the latter case resulting in their relative sliding and even bundle disintegration (Fig. 1b). In one case, two polarity-sorted bundles met with an initial anti-parallel orientation and promptly slid off each other. Subsequently, they rotated by 180° and rebound with relative orientation opposite to their original, and this time no sliding occurred (Supplementary Video 1). Overall, the active microtubule bundles almost exclusively exhibited extensile behaviour, in contrast to actin/myosin bundles, which preferentially contract¹⁴. Consequently, as we show next, gels constituted from extensile microtubule bundles are strikingly different from actin/myosin gels, which exhibit bulk contraction^{15,16}.

Increasing the concentration of extensile microtubule bundles led to the formation of percolating bundled active networks (BANs) with unique properties (Methods). For example, in contrast to classical polymer gels, which respond only passively to externally imposed stresses, BANs exhibited internally generated fluid flows (Fig. 1d, Supplementary Video 2). These highly robust spatiotemporally chaotic flows occurred throughout millimetre-sized samples and can persist for up to 24 h, limited only by the available chemical fuel (ATP) produced through the regeneration system. To gain insight into the microscopic dynamics that drive the formation of such patterns, we visualized constituent microtubule bundles at high magnification. Similar to the dilute case, we observed extension of bundles due to internal polarity sorting. However, in contrast to the dilute case, the extending microtubule bundles were now connected to a viscoelastic network. This caused them to buckle at a critical length scale¹⁷ and subsequently fracture into smaller fragments which quickly recombined with surrounding bundles of random relative polarity, generating further extension.

Thus, BAN dynamics are determined by cycles of microtubule bundles undergoing polarity sorting, extension, buckling, fracturing and subsequent recombination (Fig. 1c, Supplementary Video 3). The emergent patterns, involving thousands of microtubule bundles, are qualitatively similar to those observed in other non-equilibrium systems such as swarming bacteria or swimming spermatozoa^{18,19}. Theoretical coarse-grained models of extensile rod suspensions predict the occurrence of hydrodynamic instabilities that may underpin such large-scale collective effects^{20,21}. Being constructed from the bottom up, the properties of BANs are easily controlled and optimized. For

example, kinesin walking velocity is tuned by ATP concentration, allowing us to examine how active flows depend on the underlying microscopic dynamics (Supplementary Video 4)⁸.

To quantify fluid flow, we embedded micrometre-sized beads into the BANs (Fig. 2a, Supplementary Video 5). The beads were coated with a polymer brush to suppress their depletion-induced binding to microtubule bundles. In the absence of ATP, the tracer particles probed a passive viscoelastic network, and their mean square displacement (MSD) was sub-diffusive (Fig. 2b). At intermediate ATP concentrations, MSDs became super-diffusive at longer timescales, while remaining sub-diffusive at shorter timescales; indicating that the beads were being advected by internally generated flows. Similar effects were observed for beads suspended in active bacterial baths²². At saturating ATP concentrations, MSDs were essentially ballistic; that is, beads moved along straight paths. At longer timescales, all MSDs should become diffusive; however, this requires measuring long trajectories that are not accessible to our imaging system.

We also examined the spatial structure of internally generated fluid flow by calculating the equal-time velocity–velocity correlation function between bead pairs, $\langle V(R) \cdot V(0) \rangle$, where R is their lateral separation (Fig. 2c). The correlation functions decayed exponentially and became anti-correlated at approximately $200 \mu\text{m}$. The decay of $\langle V(R) \cdot V(0) \rangle$ taken at different ATP concentrations collapsed onto a universal curve when rescaled by the peak velocity $\langle V(0)^2 \rangle$, revealing that the spatial structure of the internally generated fluid flow is independent of kinesin velocity (Fig. 2c). For saturating ATP concentrations, the average bead velocity ($\sim 2.2 \mu\text{m s}^{-1}$) was approximately three times the maximum kinesin velocity⁸. Previous work on unbundled microtubule–motor mixtures described temporally transient dynamics before settling into a quasi-static state consisting of either aster or vortex defects¹⁰.

BANs are fundamentally different at both macroscopic and microscopic scales. At macroscopic scales, BANs exhibit steady-state permanent flows that persist for days and are only limited by ATP availability and stability of the constituent proteins. At microscopic scales, BAN dynamics is driven by cascades of microtubule bundles extending, buckling, self-fracturing and self-healing. Such unique dynamics inherently demands the presence of a depletion agent, absent in previous studies. Furthermore, BANs are very robust, exhibiting the same behaviour over a wide range of microscopic control parameters (0.4% to 10% PEG, $0.5\text{--}5 \text{ mg ml}^{-1}$ microtubules, $10 \mu\text{M}$ to 3 mM ATP, $1.5\text{--}120 \mu\text{g ml}^{-1}$ kinesin). Finally, as we show next, BANs under confinement exhibit qualitatively distinct behaviour from those of unbundled microtubule–kinesin networks²³.

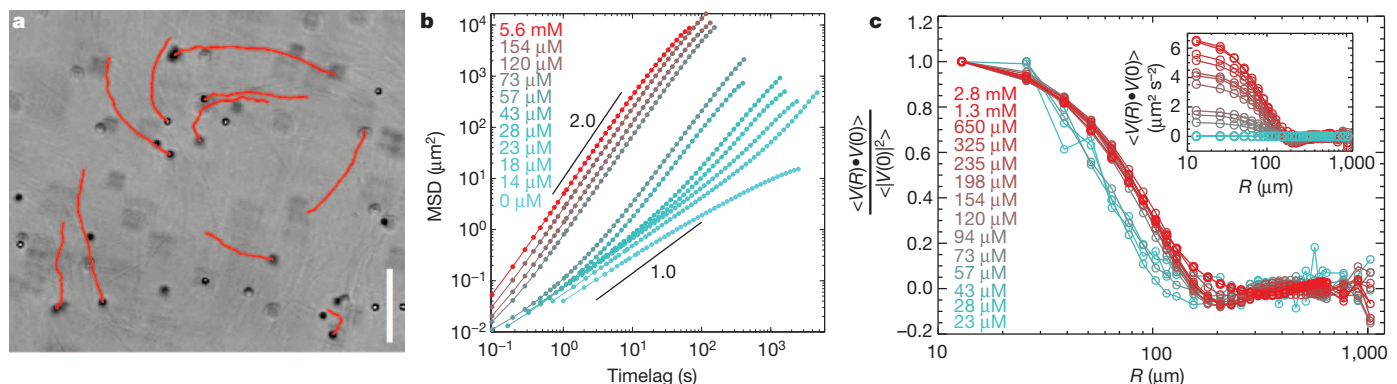


Figure 2 | ATP concentration controls dynamics of active microtubule networks. **a**, Tracer particles embedded in BANs indicate local fluid flow. Trajectories of the particles (red paths) reveal highly non-Brownian motion. Scale bar, $80 \mu\text{m}$. **b**, Mean square displacements of tracer beads plotted as a function of time for different ATP concentrations (colour-coded on the figure). Lines indicate a diffusive exponent of 1.0 and a ballistic exponent of 2.0. **c**, Normalized spatial velocity–velocity correlation functions as a function of

lateral separation for varying concentrations of ATP (colour-coded on the figure). The velocities were determined using equal time intervals of 5 s. When normalized by the peak velocity $\langle V(0)^2 \rangle$, the correlation functions rescale onto a universal curve, revealing a characteristic length scale that is independent of ATP concentration. Inset, bare spatial correlation functions reveal that average velocity depends on ATP concentration.

It is well known that with increasing concentration, rod-like molecules undergo a transition to a nematic liquid-crystalline phase. To explore this regime we created a flat 2D oil–water interface stabilized with a surfactant, with BANs being dispersed in aqueous phase. Over time extensile microtubule bundles adsorbed onto the PEG brush formed by the surfactant molecules, eventually covering the entire surface with a dense liquid-crystalline monolayer of locally aligned bundles. The adsorbed 2D layer constituted an active microtubule liquid-crystalline phase, characterized by fast streaming flows and defect unbinding (Supplementary Video 6). Further information about the nature of active microtubule liquid crystals can be garnered by examining the structure and dynamics of their defects. In general, active liquid crystals can have either nematic symmetry^{24,25} (as found in monolayers of amoeboid cells or vertically shaken monolayers of granular rods^{26,27}) or polar symmetry (which is found in motility assays at high concentrations^{5,28}). Polar liquid crystals form vortex and aster defects^{5,29}.

In contrast, we found that active microtubule liquid crystals form disclination defects of charge $1/2$ or $-1/2$, implying the presence of nematic symmetry (Fig. 3a–c). This is expected because the basic building blocks of these materials are symmetrically extensile microtubule bundles. In equilibrium liquid crystals, defects are largely static structures whose presence is determined by either internal frustrations or external boundary conditions. In contrast, defects in active microtubule nematics exhibited unique spatiotemporal dynamics. They were created as uniformly aligned nematic domains extend, buckle and internally self-fracture (Fig. 3d), similar to the dynamical cascades

of microtubule bundles in 3D active networks. Once created, a fracture line terminated with a pair of oppositely charged disclination defects. Eventually, even as the fracture self-healed, the defects remained unbound and streamed around until eventually annihilating with oppositely charged defects (Supplementary Video 6). The rates of defect creation and annihilation were balanced, creating steady-state streaming dynamics that persisted for many hours. These observations exemplify how active nematics are fundamentally different from equilibrium ones, in which fractures, internal streaming flows and spontaneous unbinding of defect pairs are never observed.

In a biological context, active fluids are frequently confined to the cytoplasm and it has been proposed that such confinement leads to emergence of coherent flows that can enhance cellular transport, a phenomenon known as cytoplasmic streaming^{30,31}. For this reason we encapsulated BANs in aqueous droplets emulsified in fluorinated oil. When squeezed between two surfaces, such water-in-oil active droplets exhibited an unforeseen emergent property: persistent autonomous motility. This motility was highly robust, limited only by the sample lifetime, which was typically a few days (Fig. 4a, Supplementary Video 7). Instead of moving along straight lines, the active emulsion droplets preferentially moved in periodic patterns, with their average velocities reaching up to $\sim 1 \mu\text{m s}^{-1}$. In comparison, droplets without any chemical fuel did not exhibit any motion (Fig. 4b).

To elucidate the microscopic mechanism that drives droplet motility, we imaged the internal microtubule dynamics. For small droplets (less than $30 \mu\text{m}$), microtubule bundles extended and pushed against

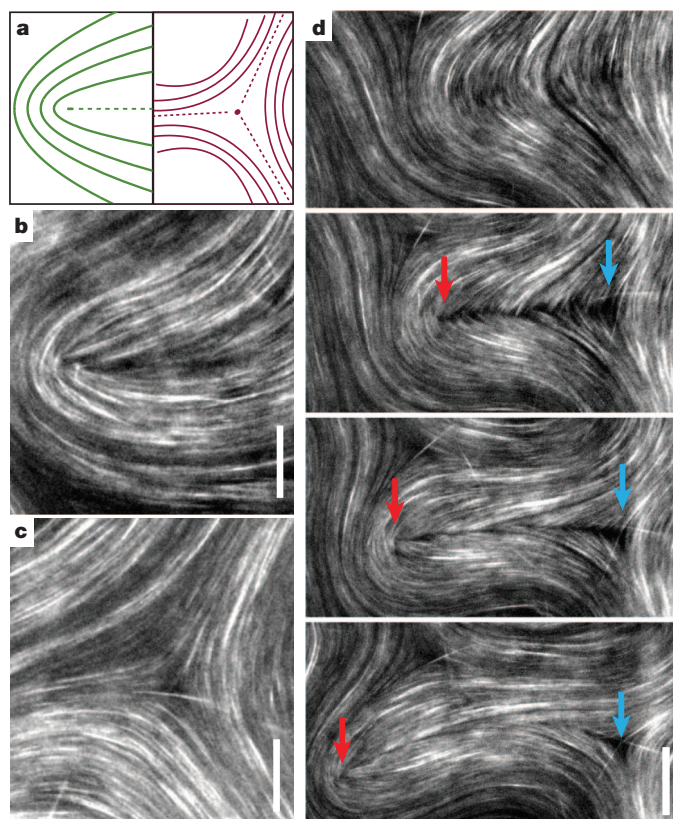


Figure 3 | Dynamics of 2D streaming nematics confined to fluid interfaces. **a**, Schematic illustrations of the nematic director configuration around disclination defects of charge $1/2$ (left) and $-1/2$ (right). **b**, **c**, Active liquid crystals exhibit disclinations of both $1/2$ (top) and $-1/2$ (bottom) charge, indicating the presence of nematic order. Scale bars, $15 \mu\text{m}$. **d**, A sequence of images demonstrates buckling, folding and internal fracture of a nematic domain. The fracture line terminates with a pair of oppositely charged disclination defects (red arrow tracks $1/2$ disclination; blue arrow tracks $-1/2$ disclination). After the fracture line self-heals, the disclination pair remains unbound. Time lapse, 15 s ; scale bar, $20 \mu\text{m}$.

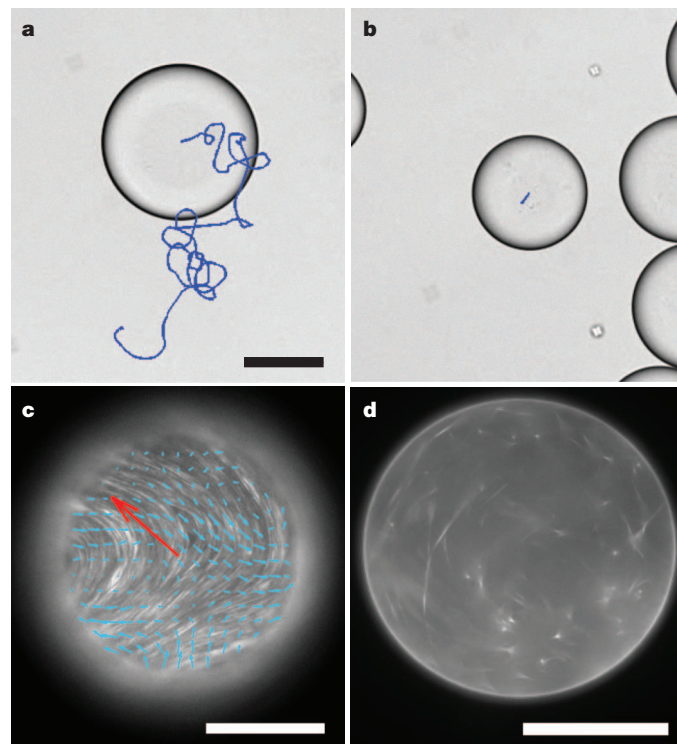


Figure 4 | Motile water-in-oil emulsion droplets. **a**, Droplets containing extensile microtubule bundles exhibit spontaneous autonomous motility when partially compressed between chamber surfaces. A droplet trajectory taken over a time interval of 33 min is overlaid onto a bright-field droplet image. Scale bar for **a** and **b** is $80 \mu\text{m}$. **b**, In the absence of ATP, passive droplets exert no internal forces, and the only contribution to their movement is minor drift. **c**, Fluorescence image of active microtubule bundles that spontaneously adsorb onto the oil–water interface. The resulting active liquid crystalline phase exhibits streaming flows, indicated with blue arrows. The red arrow indicates the direction of instantaneous droplet velocity. The image is focused on the droplet surface that is in contact with the coverslip. Scale bar, $100 \mu\text{m}$. **d**, Image of the droplet taken at a midplane indicates that the droplet interior is largely devoid of microtubule bundles. Scale bar, $100 \mu\text{m}$.

the water–oil interface, reaching a quasi-static state. Unable to buckle, they remained dispersed throughout the droplet and no motility was observed. In contrast, when confined within larger droplets, active microtubule bundles exhibited dramatically different behaviour. Similar to the flat surface, they adsorbed onto the oil–water interface, exhibiting coherent spontaneous flows (Supplementary Videos 8, 9). When in frictional contact with a hard surface, these internal flows drove the motility of the entire droplet (Fig. 4c). Scanning through the bulk of these motile droplets revealed that their interiors are largely devoid of microtubule bundles relative to their surface (Fig. 4d). The critical droplet size required for the formation of active liquid crystals was roughly similar to the collective buckling length scale that characterizes flows of 3D BANS. Spontaneous flows that drive droplet motility bear resemblance to cytoplasmic streaming observed in *Drosophila* oocytes³⁰.

Equilibrium nematics confined to spherical surfaces exhibit interesting defect configurations that have been proposed as building blocks for assembly of higher-order equilibrium structures³². Our experiments provide a new opportunity to explore how dynamic defects of active nematics on spheres can be harnessed to control droplet motility. At the next level of hierarchy, they also enable future studies in which emulsions are used for studying synchronization, crystallization and jamming of self-propelled spheres at high concentrations³³.

Conventional polymer gels, liquid crystals and emulsions are quintessential systems of soft condensed-matter physics. Studies of such materials have resulted in numerous fundamental advances and technological applications. However, the properties of these materials are limited by the laws of equilibrium statistical mechanics. Here, starting with simple extensible bundles and culminating in the assembly of motile droplets, we have demonstrated the hierarchical assembly of active gels, liquid crystals and emulsions. Liberated from equilibrium constraints, the tunable emergent properties of such far-from-equilibrium materials are strikingly different from their passive analogues, enabling further development of the statistical mechanics of active matter.

METHODS SUMMARY

Microtubules were polymerized to a concentration of 6 mg ml^{−1} with GMPCPP. The non-hydrolysing analogue of GTP reduces the microtubule nucleation barrier, resulting in very short filaments. Left at room temperature, the filaments anneal end-to-end, slowly increasing in average length. The behaviour of active microtubule networks is very sensitive to the length distribution of constituent filaments. After polymerization, microtubules used for active gel experiments were left at room temperature for two days, while microtubules used to assemble active 2D nematics were left for one. They had average lengths of 1.5 μm and 1.5 μm, respectively. Biotin-labelled kinesin fragments (401 amino acids of the amino-terminal motor domain of *Drosophila melanogaster* kinesin) were purified from *Escherichia coli*. Motor clusters were assembled by mixing biotin-labelled K401 with tetrameric streptavidin at a molar ratio of 1.7:1. Active mixtures consisted of microtubules, kinesin clusters, depleting polymer (PEG, 20 kDa), anti-bleaching agents, and an ATP regenerating system. The regenerating system used phosphoenol pyruvate (PEP) and pyruvate kinase/lactate dehydrogenase (PK/LDH) to maintain the ATP concentration at a constant level, allowing us to tune the kinesin velocity by controlling the initial ATP concentration. At the highest PEP concentrations, the regenerating system kept samples active for tens of hours. Samples were observed in a conventional flow cell and surfaces were coated with a repulsive polyacrylamide brush to suppress adsorption of proteins onto walls due to depletion or other non-specific interactions. The behaviour of active samples was monitored by visualizing microtubules with fluorescence microscopy, and by tracking 3-μm beads with bright-field microscopy. Beads were coated with poly-L-lysine-poly(ethylene glycol) (PLL-PEG) block copolymers, which suppressed their sticking to the microtubule bundles. Emulsions were made with a PFPE-PEG-PFPE surfactant (E2K0660) in HFE7500 oil.

Full Methods and any associated references are available in the online version of the paper.

Received 4 June; accepted 12 September 2012.

Published online 7 November 2012.

1. Bray, D. *Cell Movements: From Molecular to Motility* (Garland Publishing, 2000).

- Marchetti, M. C. *et al.* Soft active matter. Preprint at <http://arxiv.org/abs/1207.2929> (2012).
- Fletcher, D. A. & Geissler, P. L. Active biological materials. *Annu. Rev. Phys. Chem.* **60**, 469–486 (2009).
- MacKintosh, F. C. & Schmidt, C. F. Active cellular materials. *Curr. Opin. Cell Biol.* **22**, 29–35 (2010).
- Schaller, V., Weber, C., Semmrich, C., Frey, E. & Bausch, A. R. Polar patterns of driven filaments. *Nature* **467**, 73–77 (2010).
- Sanchez, T., Welch, D., Nicastro, D. & Dogic, Z. Cilia-like beating of active microtubule bundles. *Science* **333**, 456–459 (2011).
- Sumino, Y. *et al.* Large-scale vortex lattice emerging from collectively moving microtubules. *Nature* **483**, 448–452 (2012).
- Schnitzer, M. J. & Block, S. M. Kinesin hydrolyses one ATP per 8-nm step. *Nature* **388**, 386–390 (1997).
- Nédélec, F. J., Surrey, T., Maggs, A. C. & Leibler, S. Self-organization of microtubules and motors. *Nature* **389**, 305–308 (1997).
- Surrey, T., Nédélec, F., Leibler, S. & Karsenti, E. Physical properties determining self-organization of motors and microtubules. *Science* **292**, 1167–1171 (2001).
- Kruse, K. & Jülicher, F. Actively contracting bundles of polar filaments. *Phys. Rev. Lett.* **85**, 1778–1781 (2000).
- Liverpool, T. B. & Marchetti, M. C. Instabilities of isotropic solutions of active polar filaments. *Phys. Rev. Lett.* **90**, 138102 (2003).
- Hentrich, C. & Surrey, T. Microtubule organization by the antagonistic mitotic motors kinesin-5 and kinesin-14. *J. Cell Biol.* **189**, 465–480 (2010).
- Thoresen, T., Lenz, M. & Gardel, M. L. Reconstitution of contractile actomyosin bundles. *Biophys. J.* **100**, 2698–2705 (2011).
- Mizuno, D., Tardin, C., Schmidt, C. F. & MacKintosh, F. C. Nonequilibrium mechanics of active cytoskeletal networks. *Science* **315**, 370–373 (2007).
- Köhler, S., Schaller, V. & Bausch, A. R. Structure formation in active networks. *Nature Mater.* **10**, 462–468 (2011).
- Brangwynne, C. P. *et al.* Microtubules can bear enhanced compressive loads in living cells because of lateral reinforcement. *J. Cell Biol.* **173**, 733–741 (2006).
- Dombrowski, C., Cisneros, L., Chatkaew, S., Goldstein, R. E. & Kessler, J. O. Self-concentration and large-scale coherence in bacterial dynamics. *Phys. Rev. Lett.* **93**, 098103 (2004).
- Riedel, I. H., Kruse, K. & Howard, J. A self-organized vortex array of hydrodynamically entrained sperm cells. *Science* **309**, 300–303 (2005).
- Simha, R. A. & Ramaswamy, S. Hydrodynamic fluctuations and instabilities in ordered suspensions of self-propelled particles. *Phys. Rev. Lett.* **89**, 058101 (2002).
- Saintillan, D. & Shelley, M. J. Emergence of coherent structures and large-scale flows in motile suspensions. *J. R. Soc. Interface* **9**, 571–585 (2012).
- Wu, X. L. & Libchaber, A. Particle diffusion in a quasi-two-dimensional bacterial bath. *Phys. Rev. Lett.* **84**, 3017–3020 (2000).
- Pinot, M. *et al.* Effects of confinement on the self-organization of microtubules and motors. *Curr. Biol.* **19**, 954–960 (2009).
- Chaté, H., Ginelli, F. & Montagne, R. Simple model for active nematics: quasi-long-range order and giant fluctuations. *Phys. Rev. Lett.* **96**, 180602 (2006).
- Giomi, L., Mahadevan, L., Chakraborty, B. & Hagan, M. F. Excitable patterns in active nematics. *Phys. Rev. Lett.* **106**, 218101 (2011).
- Kemker, R., Kling, D., Kaufmann, D. & Gruler, H. Elastic properties of nematoid arrangements formed by amoeboid cells. *Eur. Phys. J. E* **1**, 215–225 (2000).
- Narayan, V., Ramaswamy, S. & Menon, N. Long-lived giant number fluctuations in a swarming granular nematic. *Science* **317**, 105–108 (2007).
- Giomi, L., Marchetti, M. C. & Liverpool, T. B. Complex spontaneous flows and concentration banding in active polar films. *Phys. Rev. Lett.* **101**, 198101 (2008).
- Kruse, K., Joanny, J. F., Jülicher, F., Prost, J. & Sekimoto, K. Asters, vortices, and rotating spirals in active gels of polar filaments. *Phys. Rev. Lett.* **92**, 078101 (2004).
- Serbus, L. R., Cha, B. J., Theurkauf, W. E. & Saxton, W. M. Dynein and the actin cytoskeleton control kinesin-driven cytoplasmic streaming in *Drosophila* oocytes. *Development* **132**, 3743–3752 (2005).
- Woodhouse, F. G. & Goldstein, R. E. Spontaneous circulation of confined active suspensions. *Phys. Rev. Lett.* **109**, 168105 (2012).
- Nelson, D. R. Toward a tetravalent chemistry of colloids. *Nano Lett.* **2**, 1125–1129 (2002).
- Henkes, S., Fily, Y. & Marchetti, M. C. Active jamming: self-propelled soft particles at high density. *Phys. Rev. E* **84**, 040301 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We acknowledge discussions with R. Meyer, R. Bruinsma and E. Frey about the nature of liquid crystalline defects. Biotin-labelled kinesin 401 (K401) was a gift from J. Gelles. This work was supported by the W. M. Keck Foundation, the National Institute of Health (5K25GM85613), the National Science Foundation (NSF-MRSEC-0820492, NSF-MRI 0923057) and the Pioneer Research Center Program through the National Research Foundation of Korea (2012-0001255). We acknowledge use of the MRSEC Optical Microscopy and Microfluidics facilities.

Author Contributions T.S., D.T.N.C., S.J.D. and Z.D. conceived the experiments and interpreted the results. T.S., D.T.N.C. and S.J.D. performed the experiments. T.S. and D.T.N.C. conducted data analysis. T.S., D.T.N.C. and Z.D. wrote the manuscript. M.H. synthesized surfactant and contributed to the assembly of active emulsions.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Z.D. (zdogg@brandeis.edu).

METHODS

Tubulin purification and microtubule polymerization. Tubulin was purified from bovine brain through two cycles of polymerization–depolymerization in high-molarity 1,4-piperazinediethanesulphonic (PIPES) buffer³⁴. The purified protein had a concentration of 7.4 mg ml⁻¹ in M2B buffer (80 mM PIPES, pH 6.8, 1 mM EGTA, 2 mM MgCl₂) and was stored at -80 °C. Tubulin was subsequently recycled and flash-frozen at a concentration of 13.2 mg ml⁻¹ in liquid nitrogen using thin-walled polymerase chain reaction (PCR) tubes. For fluorescence microscopy, tubulin was labelled with Alexa Fluor 647 (Invitrogen, A-20006) by a succinimidyl ester linker; absorbance spectroscopy indicated that 29% of monomers were labelled³⁵. Recycled tubulin was copolymerized with fluorescently labelled tubulin, producing microtubules with 3% of labelled monomers. The polymerization mixture consisted of 4.5 µl (13.2 mg ml⁻¹) unlabelled tubulin, 0.9 µl (7.4 mg ml⁻¹) Alexa-647-labelled tubulin, 1.4 µl glycerol (to 15% of final volume), 0.56 µl GMPCPP (10 mM)³⁵ (Jena Biosciences, NU-4056), 0.47 µl dithiothreitol (DTT, 20 mM), and M2B buffer to a final volume of 9.4 µl (for a final tubulin concentration of 6.8 mg ml⁻¹). The suspension was incubated for 30 min at 37 °C, and subsequently diluted to 6 mg ml⁻¹ with M2B. Microtubules were annealed at room temperature for 2 days before using them for mixing experiments. This resulted in an average microtubule length of 1.5 µm.

Kinesin–streptavidin complexes. K401, which consists of 401 amino acids of the N-terminal motor domain of *D. melanogaster* kinesin, was purified as previously published³⁶. The protein was frozen at 0.7 mg ml⁻¹ in 50 mM imidazole (pH 6.7), 4 mM MgCl₂, 2 mM DTT, 50 µM ATP and 36% sucrose buffer. Kinesin–streptavidin complexes were assembled by mixing 7 µl of freshly thawed K401 solution containing 3 mM DTT with 8 µl of 0.35 mg ml⁻¹ streptavidin (Invitrogen, S-888). The mixture was incubated on ice for at least 10 min before diluting with M2B to a final volume of 28 µl.

Polymer-coated sterically repulsive microspheres. Coating microspheres with a polymer brush leads to a steric repulsion that prevents their adsorption onto surfaces³⁷. Negatively charged polystyrene beads were coated with a block-copolymer (PLL-PEG), consisting of poly-L-lysine backbone and poly(ethylene glycol) side chains. The positively charged lysine groups electrostatically attach to the bead surface, leaving the PEG chains extending off the surface. PLL-PEG is synthesized by conjugating *N*-hydroxysuccinimidyl ester of methoxypoly(ethylene glycol) propionic acid (Laysan Bio, MPEG-SCM-20K-1g, molecular mass 20 kDa) with poly-L-lysine (Sigma, P7890-100MG, molecular mass 20 kDa) in 50 mM sodium borate buffer, pH 8.5. The concentration of lysine amino acids is 10 mM and the concentration of 20-kDa PEG molecules is 2.86 mM. This mixture is incubated at room temperature with gentle mixing for an hour and is subsequently transferred into a dialysis membrane (pore size 5 kDa). It is dialysed two times against Millipore-purified H₂O each time for more than 8 h.

PLL-PEG block-copolymer was added to 3-µm polystyrene beads decorated with sulphate groups in a low salt buffer. The final mixture containing 10 µl PLL-PEG, 100 µl HEPES (100 mM, pH 7.5), 12 µl bead stock (8% w/v) and 878 µl H₂O was incubated for 1 h, stirred gently and sonicated every 20 min. Subsequently, the beads were centrifuged for 5 min (5,000g) and resuspended in M2B buffer at ~1.3% w/v. Beads were imaged using bright-field microscopy and their positions were tracked using well-established protocols³⁸.

Observation chambers for optical microscopy. Microscope glass slides and coverslips were coated with a polyacrylamide brush to prevent non-specific adsorption of protein³⁹. Slides were first rinsed and sonicated with hot water containing 0.5% detergent, then with ethanol, and finally with 0.1 M KOH. Subsequently, the slides were soaked in a mixture of 98.5% ethanol, 1% acetic acid, and 0.5% of the silane-bonding agent 3-(trimethoxysilyl) propylmethacrylate (Acros Organics, 216551000) for 15 min. Slides were rinsed a final time and immersed in a 2% w/v aqueous solution of acrylamide. 35 µl per 100 ml of tetramethylethylenediamine (TEMED) and 70 mg per 100 ml of ammonium persulphate were added to the acrylamide solution to promote polymerization of the polyacrylamide brush that is covalently attached to the glass surfaces. Slides and coverslips were stored in a suspension of polymerized acrylamide and used for up to two weeks. Each slide was rinsed and air-dried immediately before use.

Assembly of active microtubule bundles. To study the dynamics of dilute microtubule bundles we confined a low concentration of microtubules to a quasi-2D observation chamber. In the absence of molecular motors, depletion interaction causes formation of microtubule bundles with mixed polarity. Introducing kinesin clusters induces relative sliding and eventual disintegration of bundles in which microtubules have mixed polarity while leaving intact bundles in which microtubules have the same polarity. This leads to a steady state that consists only of bundles with the same polarity. In the dilute limit (0.25 mg ml⁻¹) the average bundle length is 9 µm and the average cross-section contains approximately 10–20 microtubule bundles. Similar bundle thicknesses are measured in 3D bundled active networks.

Active microtubule networks. Several initial mixtures were prepared separately and then combined into the final solution. Individual components were dissolved and frozen separately in M2B buffer, and then thawed freshly for use. We included anti-oxidant components (listed below) and trolox (Sigma, 238813) to avoid photo-bleaching during fluorescence imaging. Anti-oxidant solution 1 (AO1) contained 15 mg ml⁻¹ glucose and 2.5 M DTT. Anti-oxidant solution 2 (AO2) contained 10 mg ml⁻¹ glucose oxidase (Sigma G2133) and 1.75 mg ml⁻¹ catalase (Sigma, C40). Second, a high-salt M2B solution (69 mM MgCl₂) was included to raise the MgCl₂. Lastly, ATP-regenerating components, including enzyme mixture pyruvate kinase/lactate dehydrogenase (PK/LDH, Sigma, P-0294) and phosphoenol pyruvate (PEP) were included in the final mixture. Because ATP is hydrolysed by kinesin activity, the PK/LDH uses PEP as a fuel source to convert ADP back into ATP at a rate that is much faster than the ATP hydrolysis. Thus constant ATP concentration is maintained throughout the experiments until all the PEP is exhausted⁴⁰.

The components described above were mixed without microtubules and in large volumes to reduce pipetting inaccuracies, while not wasting the valuable tubulin. The 'active pre-mixture' contained the following: 1.3 µl AO1, 1.3 µl AO2, 1.7 µl ATP (from stocks with different ATP concentrations), 1.7 µl PK/LDH, 2.9 µl high-MgCl₂-M2B, 3 µl of sterically repulsive beads (3 µm beads, ~1.3% w/v), 6 µl trolox (20 mM), 8 µl PEP (200 mM), 8 µl PEG (6% w/w in M2B), 4 µl kinesin–streptavidin mix, and 12.1 µl M2B buffer. To prepare the final active mixture, 5 µl of the active pre-mixture was mixed with 1 µl of the 6 mg ml⁻¹ microtubule solution, resulting in a final microtubule concentration of 1 mg ml⁻¹.

2D active microtubule nematics at oil–water interface. A 90-µm polyacrylamide flow cell was constructed as described above. HFE7500 oil with 2% (v/v) PFPE-PEG-PFPE surfactant was flowed into the cell and immediately displaced by flowing in the active microtubule mixture. The oil preferentially wets the interface, leaving a perfectly stable flat 2D oil–water interface decorated with the surfactant. The BANs were allowed to adsorb onto the oil–water interface and then imaged with fluorescence microscopy. Experiments indicate that the PEG is necessary for the adsorption of the microtubule bundles onto the surfactant monolayer and that active nematics assemble on flat as well as curved surfaces. For these experiments, 1-day-old microtubules are used, resulting in tighter nematic defects. Furthermore, the presence of surfactant is essential for the preservation of the streaming state and adsorption is only observed in active samples. In the absence of ATP, bundles do not adsorb onto the interface, instead forming a bulk rigid gel.

Light microscopy. Active microtubule networks were viewed with epi-fluorescence microscopy (Nikon Eclipse Ti microscope). AlexaFluor 647-labelled microtubules were illuminated with a 120-W metal halide light source (X-cite 120) and a fluorescent filter cube (Semrock, Cy5-4040B-NTE). Images were acquired with a cooled charge-coupled device (CCD) camera (Andor Clara). To view a large sample area, we used the motorized stage and the Ti microscope's Perfect Focus system to acquire adjacent fields of view, which were subsequently stitched together using a MatLab routine. 3-µm tracer particles were imaged with standard bright-field microscopy, and particle tracking was performed with Matlab software³⁸. Particle image velocimetry of the bundle flows was performed using the freely available Matlab-based package PIVlab (version 1.11).

Active mixtures in aqueous emulsion droplets. The active microtubule network in aqueous suspension was mixed with HFE7500 oil at a ratio of 1:5 and vortexed briefly five times to create water droplets within the oil medium. The aqueous droplets were stabilized with a 2% v/v solution of PFPE-PEG-PFPE surfactant (E2K0660)⁴¹. Glass slides and cover slips were cleaned using the protocol previously described for preparation of polyacrylamide-coated slides. The emulsion was co-flowed into the microscope chamber with extra HFE7500 oil to disperse the droplets. Within the droplets, the dynamics of microtubule bundles were viewed with fluorescence microscopy, while the motion of the droplets themselves was examined with bright-field microscopy.

34. Castoldi, M. & Popova, A. V. Purification of brain tubulin through two cycles of polymerization–depolymerization in a high-molarity buffer. *Protein Expr. Purif.* **32**, 83–88 (2003).
35. Hyman, A. et al. Preparation of modified tubulins. *Methods Enzymol.* **196**, 478–485 (1991).
36. Young, E. C., Berliner, E., Mahtani, H. K., Perez-Ramirez, B. & Gelles, J. Subunit interactions in dimeric kinesin heavy-chain derivatives that lack the kinesin rod. *J. Biol. Chem.* **270**, 3926–3931 (1995).
37. Müller, M. et al. Surface modification of PLGA microspheres. *J. Biomed. Mater. Res. A* **66A**, 55–61 (2003).
38. Crocker, J. C. & Grier, D. G. Methods of digital video microscopy for colloidal studies. *J. Colloid Interf. Sci.* **179**, 298–310 (1996).
39. Lau, A. W. C., Prasad, A. & Dogic, Z. Condensation of isolated semi-flexible filaments driven by depletion interactions. *Europhys. Lett.* **87**, 48006 (2009).
40. Huang, T. G. & Hackney, D. D. *Drosophila* kinesin minimal motor domain expressed in *Escherichia coli*—purification and kinetic characterization. *J. Biol. Chem.* **269**, 16493–16501 (1994).
41. Holtze, C. et al. Biocompatible surfactants for water-in-fluorocarbon emulsions. *Lab Chip* **8**, 1632–1639 (2008).

Little change in global drought over the past 60 years

Justin Sheffield¹, Eric F. Wood¹ & Michael L. Roderick²

Drought is expected to increase in frequency and severity in the future as a result of climate change, mainly as a consequence of decreases in regional precipitation but also because of increasing evaporation driven by global warming^{1–3}. Previous assessments of historic changes in drought over the late twentieth and early twenty-first centuries indicate that this may already be happening globally. In particular, calculations of the Palmer Drought Severity Index (PDSI) show a decrease in moisture globally since the 1970s with a commensurate increase in the area in drought that is attributed, in part, to global warming^{4,5}. The simplicity of the PDSI, which is calculated from a simple water-balance model forced by monthly precipitation and temperature data, makes it an attractive tool in large-scale drought assessments, but may give biased results in the context of climate change⁶. Here we show that the previously reported increase in global drought is overestimated because the PDSI uses a simplified model of potential evaporation⁷ that responds only to changes in temperature and thus responds incorrectly to global warming in recent decades. More realistic calculations, based on the underlying physical principles⁸ that take into account changes in available energy, humidity and wind speed, suggest that there has been little change in drought over the past 60 years. The results have implications for how we interpret the impact of global warming on the hydrological cycle and its extremes, and may help to explain why palaeoclimate drought reconstructions based on tree-ring data diverge from the PDSI-based drought record in recent years^{9,10}.

Drought is a major natural hazard that can have devastating impacts on regional agriculture, water resources and the environment, with far-reaching impacts in an increasingly globalized world¹¹. The perceived abundance of drought in the 2000s, such as long-term events in the western United States, southeast Australia and northeast China, and the recent short-term but severe events in Russia and the central United States, hint that climate change may be a forcing factor and this is only likely to get worse, given projected climate warming and precipitation changes for the twenty-first century^{1,2}. Many studies have attributed the severity and length of recent droughts to global warming^{5,12}. The Fourth Assessment Report (AR4) of the Intergovernmental Panel on Climate Change (IPCC) summarized the evidence in the following terms: “More intense and longer droughts have been observed over wider areas since the 1970s, particularly in the tropics and subtropics. Increased drying linked with higher temperatures and decreased precipitation has contributed to changes in drought”¹³. The AR4 drew heavily from historic analyses of the PDSI, which shows an increase in drought in the last few decades, regionally^{5,14} and globally⁴, that is commensurate with the increase in global temperatures. The AR4 summary has been substantially revised, however, in the recent IPCC Special Report on Extremes³ that notes the over-reliance on the PDSI and possible overestimation of the increase in regional and global drought.

The PDSI was developed originally as an agricultural monitoring tool in the United States in the 1960s¹⁵ that helped in allocating aid to stricken farmers. It is used pervasively for operational monitoring and increasingly in studies of climate change^{2,16}. Its popularity stems from

its ease of use and long history in agricultural applications. However, the PDSI has several shortcomings because of its simplicity^{3,17,18}, including the treatment of potential evaporation (PE, the evaporative demand of the atmosphere), which is calculated from temperature data by using the empirical Thornthwaite equation⁷. It has been well established that evaporation is a function of more than just temperature, and the correct physics includes radiative and aerodynamic controls on evaporative demand^{8,19–21}. Temperature-based PE methods apparently perform relatively well in climatological applications because air temperature is correlated with net radiation and humidity at weekly, monthly and subannual timescales²². However, estimating trends is problematic and there is extensive literature showing that temperature-based methods are flawed, inherently because the temperature state does not uniquely determine the evaporative flux (see Supplementary Information). In the context of climate change, the temperature-based approach responds to recent observed warming with increasing PE. When used in the PDSI model, which derives soil moisture from the balance between precipitation, evaporation and runoff, the increase in PE drives an increase in drought globally¹³ in addition to the impact of any changes in precipitation. However, numerous studies based on observations and detailed physical modelling have shown regional declines in evaporative demand over past decades as a result of various combinations of declining radiation, vapour-pressure deficit and/or wind speed⁶, despite generally increasing regional temperatures.

To resolve this discrepancy and provide an improved estimate of changes in global drought over the past 60 years based on better

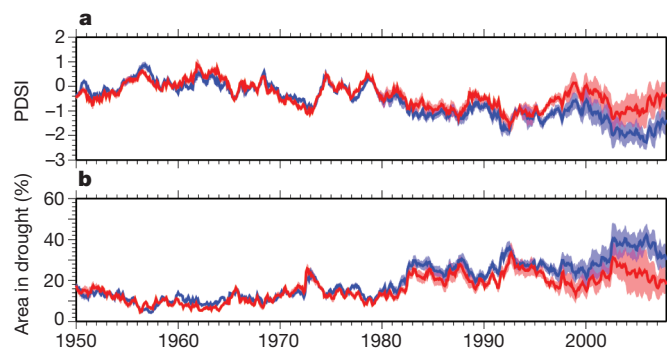


Figure 1 | Global average time series of the PDSI and area in drought. **a**, PDSI_Th (blue line) and PDSI_PM (red line). **b**, Area in drought (PDSI < -3.0) for the PDSI_Th (blue line) and PDSI_PM (red line). The shading represents the range derived from uncertainties in precipitation (PDSI_Th and PDSI_PM) and net radiation (PDSI_PM only). Uncertainty in precipitation is estimated by forcing the PDSI_Th and PDSI_PM by four alternative global precipitation data sets. Uncertainty from net radiation is estimated by forcing the PDSI_PM with a hybrid empirical–satellite data set³¹ and an empirical estimate. The other near-surface meteorological data are from a hybrid reanalysis–observational data set³¹. The thick lines are the mean values of the different PDSI data sets. The time series are averaged over global land areas excluding Greenland, Antarctica and desert regions with a mean annual precipitation of less than 0.5 mm d⁻¹.

¹Department of Civil and Environmental Engineering, Princeton University, Princeton, New Jersey 08544, USA. ²Australian Research Council Centre of Excellence for Climate System Science, Research School of Earth Science & Research School of Biology, The Australian National University, Canberra ACT 0200, Australia.

physics, we calculate global changes in the PDSI using two methods for PE. We use the Thornthwaite algorithm and a physically based estimate based on the currently accepted Penman–Monteith (PM) equation^{19,23} forced by our global meteorological data set and a set of alternative precipitation and net radiation data sets (see Methods). Because the PM is a more accurate, comprehensive and physically based model of PE (see Supplementary Information), it also has greater data requirements, which until recently has precluded its use at large scales. However, the increasing availability of detailed meteorological data from gridded ground observations, satellite remote sensing and atmospheric model reanalyses now makes it possible to calculate improved estimates that take into account radiative and aerodynamic controls. Recent studies have claimed that there is little difference between the PDSIs that use the Thornthwaite and PM algorithms (PDSI_Th and PDSI_PM, respectively)^{24,25} but this can be attributed to inconsistencies in the forcing data sets and simulation configuration (see Supplementary Information).

The global averaged time series of PDSI using the two PE methods (Fig. 1a) clearly shows the decreasing trend in the PDSI_Th since the 1970s but not for the PDSI_PM. Uncertainty due to precipitation and net radiation is estimated using four different global precipitation data sets and two different net radiation data sets, and is represented by the shading. The trend for 1950–2008 is $-0.037 \pm 0.004 \text{ yr}^{-1}$ ($P < 0.01$) and $-0.018 \pm 0.005 \text{ yr}^{-1}$ ($P < 0.01$) for the PDSI_Th and PDSI_PM, respectively. The more recent trend over 1980–2008 is similar for the PDSI_Th ($-0.032 \pm 0.008 \text{ yr}^{-1}$; $P < 0.01$) but is essentially zero for the PDSI_PM ($0.003 \pm 0.018 \text{ yr}^{-1}$; not significant). The global area in

drought (Fig. 1b) from 1980 to 2008 increases by $0.559 \pm 0.117\%$ per year ($P < 0.01$) for the PDSI_Th and increases by $0.078 \pm 0.205\%$ per year (not significant) for the PDSI_PM. Despite the smaller overall trend in PDSI_PM, there is still an increase in drought area, although this is seven times smaller than the PDSI_Th drought area trend. The spatial distribution of trends in the PDSI_Th (Fig. 2) shows drying across much of the global land, particularly over Africa and eastern Asia. In contrast, the PDSI_PM shows a mixture of drying and wetting that combines to give a smaller trend globally. The equivalent trends in the PE_Th are increasing everywhere (98% of land area), as expected given the global increase in temperature, but are a mixture of increases (58% of land area) and decreases (42%) for the PE_PM. The two methods disagree in the sign of the trends across much of northern South America, Central America, eastern North America, eastern sub-Saharan Africa, western Russia, southern and southeast Asia, and Australia. Because some of these regions are water-limited, the impact on actual evaporation, and therefore the PDSI, is small. However, in energy-limited regions such as northern Eurasia and the Amazon, the differences in PE translate into differences in the sign of the PDSI trends.

The results show that previous calculations of the increase in global drought are overestimated. However, there are several sources of uncertainty in our approach, not least from the errors in the meteorological data. We use contemporary data, which are the best that are currently available globally, but we recognize that they are not perfect. Nevertheless, the regions of decreasing PE trends estimated with the PM model are generally in agreement with the abundance of evidence

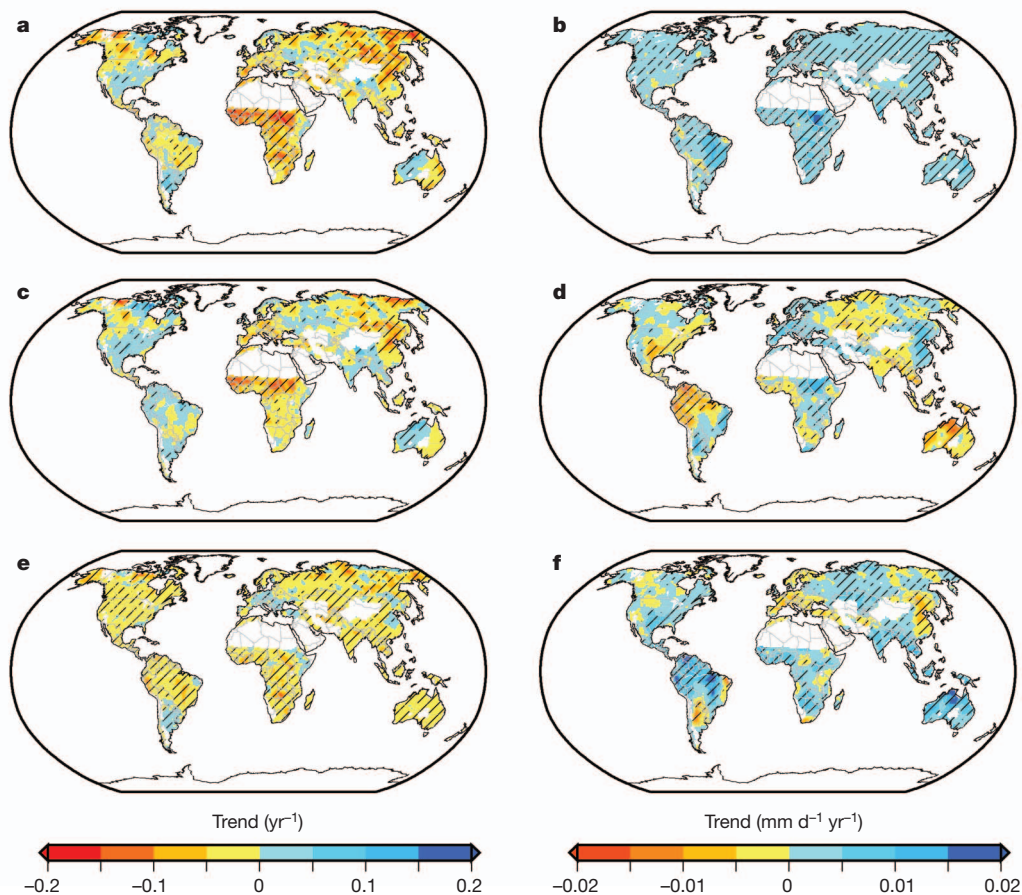


Figure 2 | Trends in the PDSI and PE. a, c, e, Non-parametric trends for 1950–2008 in annual average PDSI (averaged over the results using the four precipitation data sets and, for the PDSI_PM, also over the two net radiation data sets) from the PDSI_Th (a) and the PDSI_PM (c), and their difference (e). b, d, f, Non-parametric trends for 1950–2008 in annual average PE from the Thornthwaite equation

(b) and the PM equations (d), and their difference (f). Values are not shown for Greenland, Antarctica and desert regions with a mean annual precipitation of less than 0.5 mm d⁻¹. Statistically significant trends at the 95% level are indicated by hatching. The difference in trends in e and f and its statistical significance are calculated from the time series of differences between the two data sets.

of decreasing pan evaporation for many regions⁶, which has been attributed to global dimming, decreases in wind speed and in some places decreasing vapour-pressure deficit²⁰ (see Supplementary Information for further discussion). The temperature-based algorithm does not capture these trends. The trends in the radiation, wind-speed and humidity data used to calculate the PE_PM are generally consistent with regional observations, although increasing trends in downward longwave radiation and decreasing trends in wind speed from sparse monitoring networks are generally underestimated by the global data set. However, the possible underestimation of trends in downward longwave radiation cannot explain the difference between the PDSI_Th and PDSI_PM, and underestimation of wind-speed decline implies that the difference is probably a conservative estimate (see Supplementary Information). Furthermore, uncertainty due to the data on precipitation and net radiation as estimated from the standard deviation of PDSI trends for different precipitation and net radiation data sets is much smaller than the difference in PDSI trends due to the PE method. As these data sets improve and the uncertainties are reduced, the magnitude of the calculated trends in drought extent will change; however, the conclusion—that poor physical representation of potential evapotranspiration induces untenable estimates of long-term changes in drought—will remain.

Despite the long-standing consensus that the underlying science for temperature-based estimates of PE is flawed, compounded by the results of this and other studies^{6,21,22} that the flaws are manifested in errors in the estimations of the impact of warming on drought and hydrology in general, the reasons for the long and continued use of the PDSI_Th for climate studies in essentially its original form are a curiosity. The arguments justifying its use are generally based on the availability of data, but they are also probably related to its traditional use for mapping agricultural drought and allocating drought aid, in which comparisons—and thus division of aid—are based on spatial maps made over a short period (weeks to months). Those applications rely on spatial variations that are driven mainly by precipitation deficits, and so the PDSI_Th should do a reasonable job of distinguishing between regions of more or less drought severity at a given instant. However, in the assessment of long-term variations due to climate variability or change, a use for which the PDSI_Th was originally not designed, it seems that the over-sensitivity to changes in temperature, and other simplifications, compromises the comparisons in time. Palaeoclimate reconstructions of drought may be particularly susceptible because they are often developed by scaling tree-ring data to match the calculated PDSI_Th for their overlap period. For some regions, the tree-ring data, which reflect real variations in climatic and non-climatic factors (such as disturbances), diverge from the instrumental-based PDSI_Th in recent decades when warming has been most rapid (see, for example, refs 9, 10). Similarly, the ‘divergence problem’²⁶ as it relates to reconstructions of temperature from high-latitude and high-elevation tree-ring data may be associated with the assumption that temperature can be used as a surrogate for the controls on growth through variations in evapotranspiration, notwithstanding the competing impacts of other environmental factors (for example, higher concentrations of CO₂). This can lead to overestimation of past changes and conversely underestimation of recent trends in the context of the past.

The results of this study have implications for how we interpret the role of global warming in changes to the terrestrial hydrological cycle and its extremes, such as drought, and how we quantify the impacts of future climate change. Several regional studies^{5,12} have suggested that higher temperatures than normal were the cause for increased drought in recent years through increased evaporation. Yet there is evidence that the direct impact of temperature on drying may actually be a misinterpretation of feedbacks between the land and the atmosphere. It is more plausible that evaporation actually decreases during drought²⁷ because of less precipitation, and that drought drives increases in temperatures because there is less evaporative cooling and thus a higher sensible heat flux warming the air²⁸. Short-term temperature anomalies

are likely to be a response to drought, rather than a factor in forcing drought²⁹. Of concern is if the perceived influence of warming on drought as quantified by empirical approaches is extrapolated into the future and predictions of the impacts of climate change are likely to be overestimated^{21,22,30}. It is therefore essential to retain a perspective on the magnitude of impacts of global warming that is based on our physical understanding of the complex relationships between climate and hydrological variability. The use of physically realistic hydrological modelling merged with the wealth of *in situ* and satellite-based data sources has the potential to give better estimates of changes in global drought and its relationship with climate change.

METHODS SUMMARY

We quantify drought with the original and self-calibrating version of the PDSI model¹⁸, which uses the Thornthwaite algorithm (PDSI_Th), and a modified version (PDSI_PM) that uses the PM formulation for PE. The two models (PDSI_Th and PDSI_PM) are forced with precipitation and temperature data from our global meteorological data set³¹, which combines atmospheric reanalysis data with available remote sensing and ground observations and has been updated to 1948–2008. The meteorological data are adjusted to remove spurious trends due to observational system changes (see Supplementary Information). We also use a set of alternative global precipitation data sets to evaluate the impact of uncertainties in global precipitation trends on the drought trends, recognizing that precipitation is the main driver of drought variability but that there are uncertainties in precipitation trends at regional to global scales. The PDSI_PM model additionally requires radiation, humidity and other near-surface meteorological inputs, which are also taken from the updated meteorological data set (see Supplementary Information). Trends in annual mean values are calculated by using the non-parametric Mann–Kendall test and given as the median value across PDSI data sets derived from different precipitation data sets, with the standard deviation given after the \pm sign and estimated by the scaled median absolute deviation. An α value of 0.05 is used to test for significance. The area in drought is calculated as the percentage of land area with a PDSI of less than -3 .

Full Methods and any associated references are available in the online version of the paper.

Received 23 July; accepted 11 September 2012.

- Sheffield, J. & Wood, E. F. Projected changes in drought occurrence under future global warming from multi-model, multi-scenario, IPCC AR4 simulations. *Clim. Dyn.* **13**, 79–105 (2008).
- Dai, A. Drought under global warming: a review. *Wiley Interdisc. Rev. Clim. Change* **2**, 45–65 (2010).
- Seneviratne, S. I. *et al.* in *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation* (eds Field, C. B. *et al.*) 109–230 (Intergovernmental Panel on Climate Change, 2012).
- Dai, A., Trenberth, K. E. & Qian, T. A global data set of Palmer Drought Severity Index for 1870–2002: relationship with soil moisture and effects of surface warming. *J. Hydrometeorol.* **5**, 1117–1130 (2004).
- Briffa, K. R., van der Schrier, G. & Jones, P. D. Wet and dry summers in Europe since 1750: evidence of increasing drought. *Int. J. Climatol.* **29**, 1894–1905 (2009).
- Roderick, M. L., Hobbins, M. T. & Farquhar, G. D. Pan evaporation trends and the terrestrial water balance II. Energy balance and interpretation. *Geog. Compass* **3**, 761–780 (2009).
- Thornthwaite, C. W. An approach toward a rational classification of climate. *Geogr. Rev.* **38**, 55–94 (1948).
- Penman, H. L. Natural evaporation from open water, bare soil, and grass. *Proc. R. Soc. Lond. A* **193**, 120–145 (1948).
- Fang, K. Y. *et al.* Drought variations in the eastern part of northwest China over the past two centuries: evidence from tree rings. *Clim. Res.* **38**, 129–135 (2009).
- de Grandpré, L. *et al.* Seasonal shift in the climate responses of *Pinus sibirica*, *Pinus sylvestris*, and *Larix sibirica* trees from semi-arid, north-central Mongolia. *Can. J. For. Res.* **41**, 1242–1255 (2011).
- Sternberg, T. Regional drought has a global impact. *Nature* **472**, 169 (2011).
- Cai, W., Cowan, T., Briggs, P. & Raupach, M. Rising temperature depletes soil moisture and exacerbates severe drought conditions across southeast Australia. *Geophys. Res. Lett.* **36**, L21709 (2009).
- IPCC. *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Solomon, S. *et al.*) (Cambridge Univ. Press, 2007).
- Wang, J., Chen, F., Jin, L. & Bai, H. Characteristics of the dry/wet trend over arid central Asia over the past 100 years. *Clim. Res.* **41**, 51–59 (2010).
- Palmer, W. C. *Meteorological Drought* (US Department of Commerce Research Paper 45, 1965).
- Zhao, M. & Running, S. W. Drought-induced reduction in global terrestrial net primary production from 2000 through 2009. *Science* **329**, 940–943 (2010).

17. Alley, W. M. The Palmer Drought Severity Index: limitations and assumptions. *J. Clim. Appl. Meteorol.* **23**, 1100–1109 (1984).
18. Wells, N., Goddard, S. & Hayes, M. J. A self-calibrating Palmer drought severity index. *J. Clim.* **17**, 2335–2351 (2004).
19. Shuttleworth, W. J. in *Handbook of Hydrology* (ed. Maidment, D. R.) 4.1–4.53 (McGraw-Hill, 1993).
20. Roderick, M. L., Rotstain, L. D., Farquhar, G. D. & Hobbins, M. T. On the attribution of changing pan evaporation. *Geophys. Res. Lett.* **34**, L17403 (2007).
21. Donohue, R. J., McVicar, T. R. & Roderick, M. L. Assessing the ability of potential evaporation formulations to capture the dynamics in evaporative demand within a changing climate. *J. Hydrol. (Amst.)* **386**, 186–197 (2010).
22. Shaw, S. & Riha, S. J. Assessing temperature-based PET equations under a changing climate in temperate, deciduous forests. *Hydrol. Process.* **25**, 1466–1478 (2011).
23. Monteith, J. L. Evaporation and environment. *Symp. Soc. Exp. Biol.* **19**, 205–234 (1964).
24. Dai, A. Characteristics and trends in various forms of the Palmer Drought Severity Index during 1900–2008. *J. Geophys. Res.* **116**, D12115 (2011).
25. van der Schrier, G., Jones, P. D. & Briffa, K. R. The sensitivity of the PDSI to the Thornthwaite and Penman–Monteith parameterizations for potential evapotranspiration. *J. Geophys. Res.* **116**, D03106 (2011).
26. D'Arrigo, R., Wilson, R., Liepert, B. & Cherubini, P. On the 'Divergence Problem' in northern forests: a review of the tree-ring evidence and possible causes. *Global Planet. Change* **60**, 289–305 (2008).
27. Roderick, M. L. & Farquhar, G. D. Changes in Australian pan evaporation from 1970 to 2002. *Int. J. Climatol.* **24**, 1077–1090 (2004).
28. Lockart, N., Kavetski, D. & Franks, S. W. On the recent warming in the Murray–Darling Basin: land surface interactions misunderstood. *Geophys. Res. Lett.* **36**, L24405 (2009).
29. Hirschi, M. *et al.* Observational evidence for soil-moisture impact on hot extremes in southeastern Europe. *Nature Geosci.* **4**, 17–21 (2011).
30. Milly, P. C. D. & Dunne, K. A. On the hydrologic adjustment of climate-model projections: the potential pitfall of potential evapotranspiration. *Earth Interact.* **15**, 1–14 (2011).
31. Sheffield, J., Goteti, G. & Wood, E. F. Development of a 50-yr high-resolution global dataset of meteorological forcings for land surface modeling. *J. Clim.* **19**, 3088–3111 (2006).

Supplementary Information is available in the online version of the paper.

Acknowledgements J.S. acknowledges support from the US National Oceanic and Atmospheric Agency (NA100AR4310130, NA11OAR4310097) and NASA (NNX08AN40A). M.L.R. acknowledges the support of the Australian Research Council (DP0879763, DP110105376, CE11E0098).

Author Contributions J.S. and E.F.W. conceived the study with inspiration from M.L.R. J.S. performed the analyses and mainly wrote the manuscript. E.F.W. and M.L.R. contributed to discussion and the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.S. (justin@princeton.edu).

METHODS

Palmer Drought Severity Index. Historically, the PDSI¹⁵ has been the tool of choice when monitoring and analysing drought occurrence, especially in the United States, where it is one component of the US National Drought Monitor³². It is generally calculated on weekly to monthly timescales and uses precipitation and temperature inputs to drive a simplified water-balance model with a generic two-layer soil model and meteorology that is normalized with a reference set of water balance terms. At continental to global scales its simplicity makes it an attractive choice for reconstructing drought records^{4,33}, for which it has also been shown to be a proxy for soil moisture⁴. It has been used to analyse continental-scale to global-scale, long-term variability in drought by several studies (see, for example, refs 4, 34, 35). The PDSI is generally well correlated with output from more comprehensive hydrological modelling³⁶ but diverges in cooler seasons and high latitudes, and substantially so in drier regions.

Despite this legacy, the PDSI has been shown to be unsuitable for widespread application and suffers from simplifications in its physical basis and soil hydrology^{17,37}. For example, its exclusion of cold-season processes makes it unsuitable for application in many parts of the world. The usual form of the PDSI algorithm uses several empirical constants to characterize the local climate. These were originally derived by Palmer using data from a number of climate divisions generally located in the midwestern United States and are therefore not representative of the whole globe. For example, the criteria for signalling the start and end of a drought are arbitrarily based on original data from the midwestern United States. Some of the shortcomings have been addressed by a self-calibrating version¹⁸ that removes the spatial inconsistencies.

We quantify drought with the original and self-calibrating version of the PDSI model¹⁸, which uses the Thornthwaite algorithm (PDSI_Th), and a modified version (PDSI_PM) that uses the PM formulation for PE.

Thornthwaite PE algorithm. PE is modelled in the PDSI by using the temperature-only-based Thornthwaite method. Thornthwaite⁷ correlated mean monthly temperature with PE, as determined from the water balance for valleys in the eastern USA, where sufficient moisture was available to maintain active transpiration. The Thornthwaite formula for monthly PE (mm) is

$$PE = 16d(10T/I)^a$$

where T is the mean temperature for the month (in °C) and d is a correction factor that depends on latitude and month. I is the annual thermal index,

$$I = \sum_{i=1}^{12} (T_i/5)^{1.514}$$

where the subscript i refers to the month of the year and a is an empirical factor,

$$a = 0.49 + 0.0179I - 0.0000771I^2 + 0.000000675I^3$$

Penman–Monteith PE algorithm. The PM approach²³ is generally accepted as the most comprehensive algorithm for modelling potential and actual evapotranspiration (given additional estimates of the plant and environmental resistance to atmospheric demand). It is derived from consideration of the equations of the surface energy balance by means of elimination of the surface temperature term. It forms the basis for the evaporation submodel of many distributed hydrological and land surface models, the latter of which often form the land component of coupled climate models, and has been used as the basis for regional and global retrievals of evapotranspiration based on satellite remotely sensed data (see, for example, refs 38–40). The PM equation given below models the diffusion of energy from plants or soil against stomatal and aerodynamic resistance, given inputs of net radiation, temperature, humidity and wind speed:

$$ET = \frac{\Delta R_{\text{net}} + (\rho c_p D / r_a)}{\Delta + \gamma(1 + r_s / r_a)}$$

where evapotranspiration (ET) is now in W m^{-2} , which can be converted into mm per month by dividing by the latent heat of vaporization of water, λ (J kg^{-1}). Δ (Pa K^{-1}) is the slope of the plot of saturated vapour pressure against air temperature, R_{net} is the net radiation (W m^{-2}), ρ is the density of air (kg m^{-3}), c_p is the specific heat of air at constant pressure ($\text{J kg}^{-1} \text{K}^{-1}$), D is the vapour-pressure deficit (Pa) and γ is the psychrometric constant (Pa K^{-1}). r_a and r_s are the aerodynamic and stomatal resistances (s m^{-1}), respectively. ET collapses to PE when the stomatal resistance is zero; the recommended form of the equation¹⁹, given as the sum of the radiative and aerodynamic components, is

$$PE = \frac{\Delta}{\Delta + \gamma} R_{\text{net}} + \frac{\gamma}{\Delta + \gamma} 6.43(1 + 0.536U)D$$

where PE is now in mm d^{-1} ; U is the wind speed (m s^{-1}) at 2 m height. In comparison with the Thornthwaite expression, which is based solely on temperature, the PM models evaporation as the combination of radiative and aerodynamic processes, thus giving a more realistic estimate and having the potential to be influenced by changes in humidity, radiation and wind speed, as well as temperature.

Global meteorological forcing dataset. The two PDSI models (PDSI_Th and PDSI_PM) are forced with precipitation and temperature data from our global meteorological data set. The PDSI_PM model additionally requires radiation, humidity and other near-surface meteorological inputs, which are also taken from this data set. The global meteorological forcing dataset³¹ combines reanalysis data and observations to form a global, long-term (1948–2008), 1.0°, 3-hourly data set of precipitation, surface radiation and near-surface meteorology. The data set is designed for forcing land-surface hydrological models and other physical models at large spatial (regional to global) and temporal (annual to decadal) domains. Thus the goal of the data set is to ensure robustness of long-term trends and variability. At the same time it adjusts the short-term (daily and diurnal) variations to match observational data where available and maintains interrelations between variables. The data set is based on the NCEP/NCAR reanalysis (NNR)⁴¹, which provides continuous records of atmospheric and land variables from 1948 to the present. The reanalysis data are corrected to remove biases at diurnal to annual timescales by merging with observational data for precipitation, air temperature, and shortwave and longwave radiation. Precipitation and temperature are scaled to match the Climatic Research Unit (CRU) TS3.0 data set⁴² on a monthly timescale. The diurnal cycle of precipitation is resampled from a statistical model derived from the Tropical Rainfall Measurement Mission (TRMM) Multi-Satellite Precipitation Analysis (TMPA) satellite-based data set⁴³. The diurnal range in air temperature is also adjusted to match the CRU TS3.0 data set. For 1984–2007, the shortwave and longwave radiation are scaled to match the NASA/GEWEX Surface Radiation Budget (SRB) satellite based data set⁴⁴ on a monthly timescale. The data set is available from <http://hydrology.princeton.edu/data.pgpf.php>.

Alternative precipitation data sets. A set of four global precipitation data sets is used to quantify the uncertainty in global precipitation. The data sets are CPC-Prec/L⁴⁵ (<ftp://ftp.cpc.ncep.noaa.gov/precip/50yr/>), Global Precipitation Climatology Centre V4 (ref. 46) (GPCC; <http://gpcc.dwd.de>), Climatic Research Unit TS3.0 (ref. 42) (<http://badc.nerc.ac.uk/browse/badc/cru>) and Willmott–Matsuura V2.01 (ref. 47) (http://climate.geog.udel.edu/~climate/html_pages/download.html).

Trend calculation. Trends in annual mean values are calculated by using the non-parametric Mann–Kendall test and given as the median value across PDSI data sets derived from different precipitation data sets, with the standard deviation given after the \pm sign and estimated by the scaled median absolute deviation. An α value of 0.05 is used to test for significance. The area in drought is calculated as the percentage of land area with a PDSI of less than -3 .

32. Svoboda, M. *et al.* The Drought Monitor. *Bull. Am. Meteorol. Soc.* **83**, 1181–1190 (2002).
33. Cook, E. R., Meko, D. M., Stahle, D. W. & Cleaveland, M. K. Drought reconstructions for the continental United States. *J. Clim.* **12**, 1145–1162 (1999).
34. van der Schrier, G., Briffa, K. R., Osborn, T. J. & Cook, E. R. Summer moisture availability across North America. *J. Geophys. Res.* **111**, D11102 (2006).
35. McCabe, G. J. & Palecki, M. A. Multidecadal climate variability of global lands and oceans. *Int. J. Climatol.* **26**, 849–865 (2006).
36. Sheffield, J. & Wood, E. F. Characteristics of global and regional drought, 1950–2000: analysis of soil moisture data from off-line simulation of the terrestrial hydrologic cycle. *J. Geophys. Res.* **112**, D17115 (2007).
37. Jensen, M. E. *Consumptive Use of Water and Irrigation Water Requirements* (American Society of Civil Engineers, 1973).
38. Mu, Q. *et al.* Development of a global evapotranspiration algorithm based on MODIS and global meteorology data. *J. Geophys. Res.* **112**, G01012 (2007).
39. Cleugh, H. A., Leuning, R., Mu, Q. & Running, S. W. Regional evaporation estimates from flux tower and MODIS satellite data. *Remote Sens. Environ.* **106**, 285–304 (2007).
40. Sheffield, J., Wood, E. F. & Munoz-Arriola, F. Long-term regional estimates of evapotranspiration for Mexico based on downscaled ISCCP data. *J. Hydrometeorol.* **11**, 253–275 (2010).
41. Kalnay, E. *et al.* The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Am. Meteorol. Soc.* **77**, 437–471 (1996).
42. Mitchell, T. D. & Jones, P. D. An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *Int. J. Climatol.* **25**, 693–712 (2005).
43. Huffman, G. J. *et al.* The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *J. Hydrometeorol.* **8**, 38–55 (2007).

44. Gupta, S. K., Stackhouse, P. W., Cox, S. J., Mikovitz, J. C. & Zhang, T. Surface Radiation Budget Project completes 22-year data set. *GEWEX News* **16**, 12–13 (2006).
45. Chen, M., Xie, P., Janowiak, J. E. & Arkin, P. A. Global land precipitation: a 50-yr monthly analysis based on gauge observations. *J. Hydrometeorol.* **3**, 249–266 (2002).
46. Schneider, U., Fuchs, T., Meyer-Christoffer, A. & Rudolf, B. Global precipitation analysis products of the GPCC. *Weather and Climate—Deutscher Wetterdienst—Klimadatenzentrum-WZN* <ftp://ftp-anon.dwd.de/pub/data/gpcc/PDF/GPCC_intro_products_2008.pdf> (2008).
47. Willmott, C. J. & Matsuura, K. Terrestrial Air Temperature: 1900–2008 Gridded Monthly Time Series, version 2.01. *Global Air Temperature Archive* <http://climate.geog.udel.edu/~climate/html_pages/Global2_Ts_2009/README.global_t_ts_2009.html> (2010).

Slowdown of the Walker circulation driven by tropical Indo-Pacific warming

Hiroki Tokinaga¹, Shang-Ping Xie^{1,2}, Clara Deser³, Yu Kosaka⁴ & Yuko M. Okumura⁵

Global mean sea surface temperature (SST) has risen steadily over the past century^{1,2}, but the overall pattern contains extensive and often uncertain spatial variations, with potentially important effects on regional precipitation^{3,4}. Observations suggest a slowdown of the zonal atmospheric overturning circulation above the tropical Pacific Ocean (the Walker circulation) over the twentieth century^{1,5}. Although this change has been attributed to a muted hydrological cycle forced by global warming^{5,6}, the effect of SST warming patterns has not been explored and quantified^{1,7,8}. Here we perform experiments using an atmospheric model, and find that SST warming patterns are the main cause of the weakened Walker circulation over the past six decades (1950–2009). The SST trend reconstructed from bucket-sampled SST and night-time marine surface air temperature features a reduced zonal gradient in the tropical Indo-Pacific Ocean, a change consistent with subsurface temperature observations⁸. Model experiments with this trend pattern robustly simulate the observed changes, including the Walker circulation slowdown and the eastward shift of atmospheric convection from the Indonesian maritime continent to the central tropical Pacific. Our results cannot establish whether the observed changes are due to natural variability or anthropogenic global warming, but they do show that the observed slowdown in the Walker circulation is presumably driven by oceanic rather than atmospheric processes.

The Walker circulation is a defining feature of tropical climate. On seasonal-to-interannual timescales, it is closely tied to the east–west SST gradient along the Equator, as is the El Niño–Southern Oscillation (ENSO). Long-term changes in the Walker circulation have recently been the subject of intense debate^{1,5–12}. Most climate models in the Coupled Model Intercomparison Project phase 3 (CMIP3) predict a slowdown of the Walker circulation under global warming^{5–7,9,13}, consistent with sea level pressure (SLP) measurements during the twentieth century^{1,5,8}. The Walker circulation weakening relaxes prevailing easterly trade winds over the equatorial Pacific^{5,8} and suppresses the upwelling of nutrient-rich cold water⁷, affecting fishery stocks in the region. Despite these important climatic and biogeochemical influences, the mechanism for the observed long-term change in the Walker circulation has not been rigorously investigated.

One major hypothesis attributes the Walker circulation slowdown to a decrease in convective mass flux under global warming that balances a slower increase in global mean precipitation than can be expected from the increase in atmospheric water vapour⁶. This muted hydrological cycle mechanism relies only on a globally uniform increase in SST³. Overlooked, however, is an alternative hypothesis that points to changes in zonal SST gradient across the tropical oceans^{1,11,14}. This mechanism is dominant for ENSO, but the effort to test its role in the long-term weakening of the Walker circulation has been hampered by uncertainty in observed SST warming patterns^{1,7,8}. For 1950–2009, the zonal SST gradient across the Indo-Pacific has intensified in the HadISST1 (Hadley Centre sea ice and SST

version 1; ref. 15) data set but shows little change in the ERSST3b (extended reconstructed SST version 3b; ref. 16) data set (Fig. 1). This disagreement is due to the change in SST measurement technique and different analysis methods⁷.

We reconstruct the SST trend pattern for the period 1950–2009 (Fig. 1a), for which various data sets are available with improved quality. To reduce biases caused by changes in SST measurement technique, we construct an SST data set using only bucket measurements from the International Comprehensive Ocean–Atmosphere Data Set¹⁷ (ICOADS). The resulting trend pattern exhibits a reduced zonal gradient between the eastern Pacific and the western Pacific/southeastern Indian Ocean (Supplementary Fig. 1a), in contrast to HadISST1 and ERSST3b results (Fig. 1b, c). We also examine night-time marine surface air temperature (NMAT), a meteorological parameter that is highly correlated with SST (Supplementary Fig. 2) but free of the biases introduced by changes in SST measurement technique. The NMAT trend pattern bears a striking resemblance to that of bucket SST (Supplementary Fig. 1b, c). This motivates us to estimate trends in merged surface temperature (MST) of the bucket SST and NMAT (see Methods). The MST trend features a reduced zonal gradient in the tropical Indo-Pacific (Fig. 1a, d), physically consistent with a flattened ocean thermocline observed by expendable bathythermographs⁸.

To investigate the SST gradient effect, we use four atmospheric general circulation models (AGCMs; see Methods) and construct a multi-model ensemble for an unbiased estimate of the atmospheric response. Each AGCM is forced with four different tropical SST trend patterns: a spatially uniform SST increase (SUSI; $\sim 0.5^\circ\text{C}$ for 60 yr from 1950 to 2009) and those estimated from MST, HadISST1 and ERSST3b. In all the experiments including SUSI, the greenhouse-gas increase from the 1950s to the 2000s is imposed. The AGCM response to SUSI is not an accurate proxy for anthropogenic change, but serves as a useful reference to evaluate the SST pattern effect. To validate the model experiments, we use a wide variety of data sets, all of which have successfully captured regional patterns of tropical climate change^{1,8} (Methods). We also use 71 ensemble members of the ‘climate of the twentieth century’ experiments (20C3M) from 24 CMIP3 models and three atmospheric reanalyses for comparison: the US National Centers for Environmental Prediction (NCEP)–National Center for Atmospheric Research (NCAR) reanalysis¹⁸, the European Centre for Medium-Range Weather Forecasting 40-year reanalysis (ERA-40; ref. 19), and the Twentieth Century Reanalysis version 2 (20CRv2; ref. 20).

Over the past six decades, a weakening trend in the Walker circulation is apparent in SLP change^{1,8,9} (Fig. 1a and Supplementary Fig. 3a), increasing over the maritime continent and decreasing east of the Philippines and over the central to eastern tropical Pacific. The MST-forced experiments reproduce these SLP change patterns quite well, with a slight eastward displacement of negative values over the eastern Pacific (Fig. 1d). In contrast, the HadISST1-forced experiments strengthen the Walker circulation¹² (Fig. 1b), while the ERSST3b experiments show no significant change (Fig. 1c). The SUSI-forced

¹International Pacific Research Center, Department of Meteorology, SOEST, University of Hawaii at Manoa, 1680 East West Road, Honolulu, Hawaii 96822, USA. ²Physical Oceanography Laboratory, Ocean University of China, Qingdao 266003, China. ³National Center for Atmospheric Research, P.O. Box 3000, Boulder, Colorado 80307, USA. ⁴International Pacific Research Center, SOEST, University of Hawaii at Manoa, 1680 East West Road, Honolulu, Hawaii 96822, USA. ⁵Institute for Geophysics, The University of Texas at Austin, 10100 Burnet Road, Austin, Texas 78758, USA.

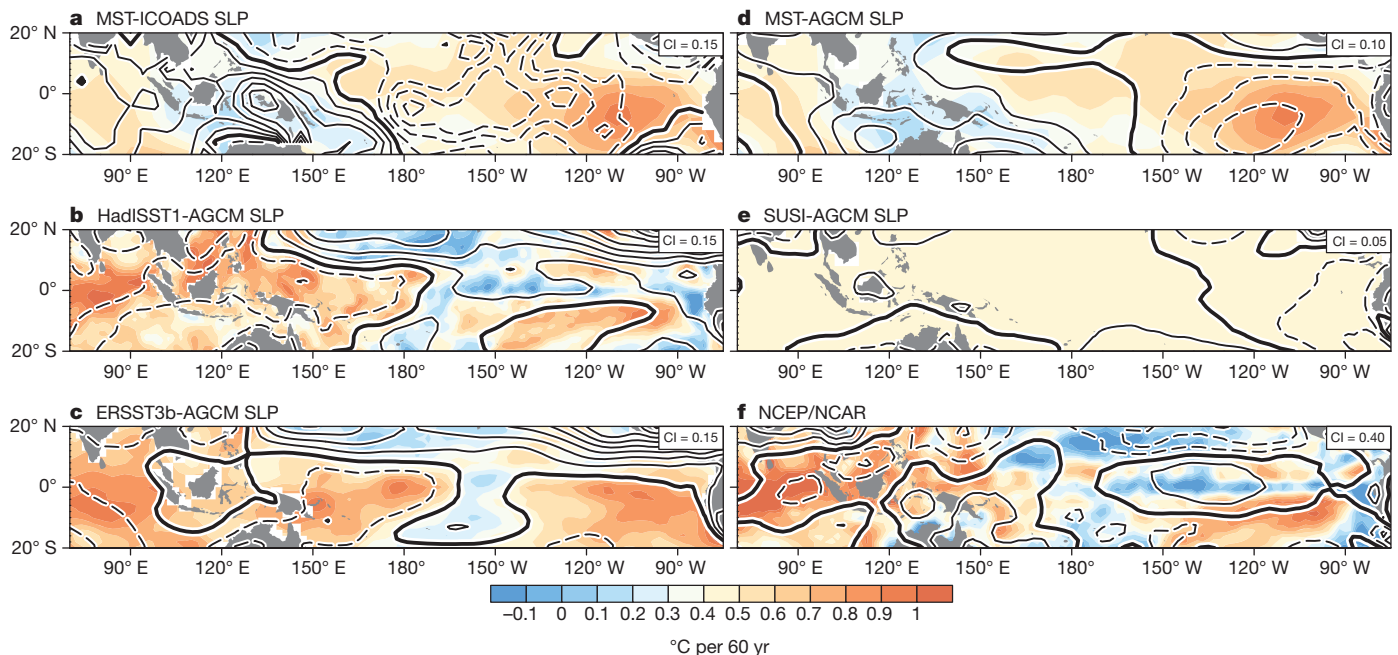


Figure 1 | Observed and simulated patterns of SLP and SST changes for 1950–2009. Changes in SLP (hPa per 60 yr; contours) and SST (°C per 60 yr; shading) from: **a**, ICOADS and MST; **b**, HadISST1-forced experiment; **c**, ERSST3b-forced experiment; **d**, MST-forced experiment; **e**, SUSI-forced

experiment; and **f**, NCEP-NCAR reanalysis. The contour interval (CI) for SLP change is indicated at the top-right corner of each panel. Positive (negative) contours are solid (dashed) lines, and zero contours are thickened. The basin averages (40°E–70°W, 20°S–20°N) of SLP change are removed.

experiments simulate a slight weakening of the Walker circulation (Fig. 1e) but its magnitude is much weaker than that of observations⁹ and MST-forced experiments. These AGCM experiments suggest that SST trend patterns are key to the Walker circulation change.

Both the NCEP-NCAR (Fig. 1f) and ERA-40 (Supplementary Fig. 3b) reanalyses show an intensification of the Walker circulation with SLP increasing over the central equatorial Pacific and decreasing over the maritime continent. The SST data sets used in these reanalyses exhibit an intensified zonal gradient similar to that in HadISST1, suggesting a strong impact of SST forcing even in the NCEP-NCAR and ERA-40 reanalyses. The 20CRv2 reanalysis exhibits no significant change in SLP gradient despite the assimilation of SLP observations (Supplementary Fig. 3c). HadISST1 is used as a surface forcing in 20CRv2, which we speculate acts to oppose the observed reduction in the Walker circulation.

Changes in atmospheric convection are a clue to the Walker circulation slowdown. Observations indicate an eastward shift of precipitation/cloudiness from the maritime continent to the central tropical Pacific, accompanied by consistent changes in surface wind convergence (Fig. 2a, b). The MST-forced experiments simulate these patterns of cloud, precipitation and surface wind changes quite well, albeit at reduced magnitudes (Fig. 2c, d). The simulated convection changes closely follow spatial patterns of SST warming, consistent with the “warmer-get-wetter” idea, with support from model projections³ and satellite observations for the past three decades²¹. In contrast, the SUSI experiments simulate a precipitation/cloudiness increase over the maritime continent (Fig. 2e, f), in disagreement with observations. Within the framework of our AGCM experiments, SST warming patterns, rather than the hydrological cycle response to uniform warming, is the main driver for the Walker circulation slowdown and atmospheric convection changes over the past six decades.

The close linkage between SST and the Walker circulation is readily illustrated in a scatter plot of changes in zonal SST and SLP gradients (denoted Δ SST and Δ SLP, respectively), defined as the zonal difference on the Equator between the eastern Pacific (150°–90°W, 5°S–5°N) and western Pacific/eastern Indian Ocean (90°–150°E, 5°S–5°N) (Fig. 3). Changes in Δ SST and Δ SLP are linearly correlated among

AGCM experiments and reanalyses ($r \approx -0.97$, significant at $P < 0.01$), with a strong dependence on SST data sets. MST-forced (HadISST1-forced) AGCMs robustly simulate a weakening (strengthening) of Δ SLP in response to the Δ SST weakening (strengthening), whereas ERSST3b-forced AGCMs show little change in Δ SLP. SUSI-forced AGCMs tend to weaken Δ SLP, but this is not robust among models and the amplitude is small. Overall, MST-forced AGCMs simulate the observed Δ SLP change more realistically than other SST data sets. In addition, Δ SLP and Δ SST are highly correlated among the 71 simulations in the CMIP3 20C3M ensemble ($r \approx -0.83$, significant at $P < 0.01$), although the Δ SLP reduction is smaller than in observations⁹.

We estimate uncertainty in SST trends arising from observational biases based on 100 realizations of SST data in the third version of the Hadley Centre gridded SST data set (HadSST3; refs 2, 22). There are large variations in Δ SST among different realizations (Figs 3 and 4a). We apply empirical orthogonal function (EOF) analysis to the 1950–2006 trend patterns over the tropical Indo-Pacific (30°E–60°W, 32.5°S–32.5°N) from the 100 realizations. The leading EOF reveals large uncertainty in warming magnitude over the equatorial Indo-western Pacific, accounting for 83.6% of the total variance among the 100 SST realizations (Supplementary Fig. 4a). The correlation between the principal component of the leading EOF and the Δ SST trend is -0.75 (significant at $P < 0.01$), indicating that the Δ SST trend is strongly affected by uncertainty in the magnitude of SST warming over the equatorial Indo-western Pacific (Fig. 4a). Nevertheless, all the HadSST3 realizations show a reduction in Δ SST. AGCM experiments forced by a subset of HadSST3 realizations, including the two extremes (no. 12 and no. 34) and median, all simulate the Walker circulation slowdown (Fig. 4b–e), with SLP increasing (decreasing) over the maritime continent (tropical Pacific). The magnitude of circulation change is proportional to that of Δ SST (green triangles in Fig. 3). These HadSST3 experiments provide strong support for our results from observations and MST-forced AGCMs.

Our AGCM results show that the observed Walker circulation slowdown is due to the weakened zonal SST gradient across the equatorial

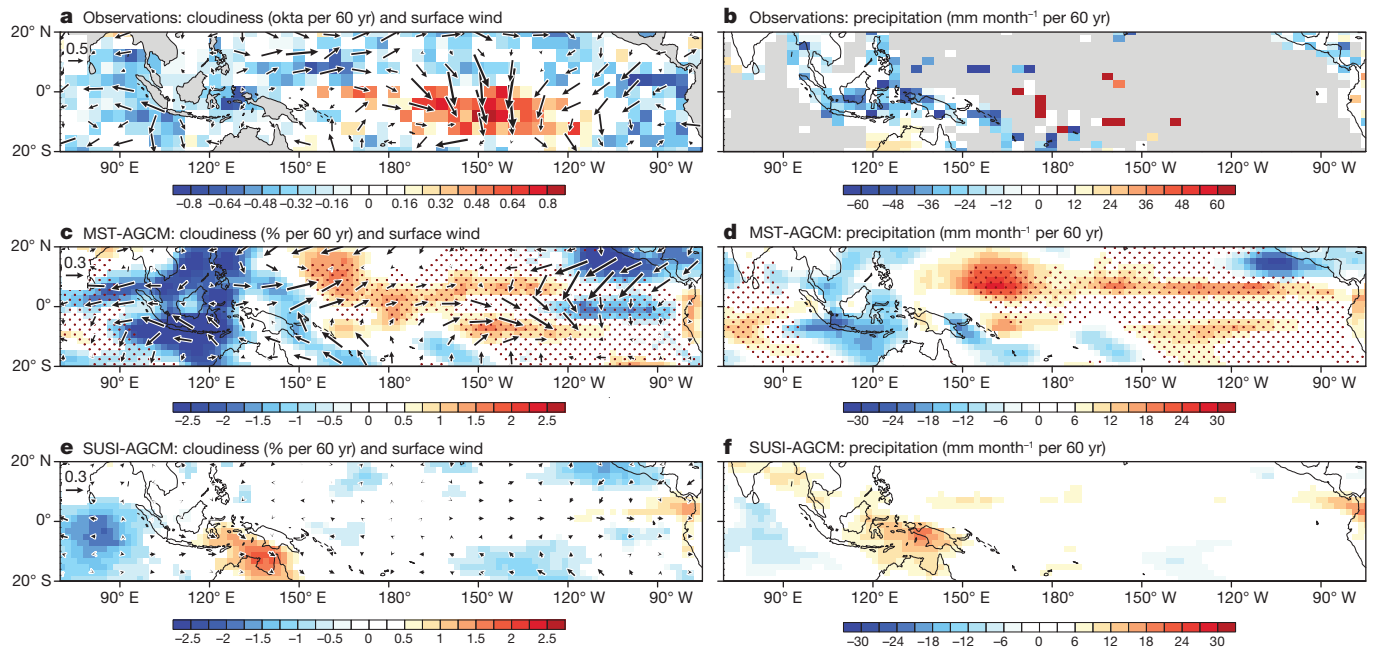


Figure 2 | Observed and simulated changes in cloudiness, surface wind and precipitation for 1950–2009. Changes in cloudiness (okta per 60 yr for observation, % per 60 yr for AGCM; shading in left panels), surface wind (m s^{-1} per 60 yr; vectors in left panels), and precipitation (mm month^{-1} per 60 yr; shading in right panels) from: **a, b**, observations (ICOADS, WASWind and

rain-gauges); **c, d**, MST-forced experiment; and **e, f**, SUSI-forced experiment. Only significant changes exceeding the 95% confidence level are shaded in colour. Stippling in **c, d**, indicates regions of the MST warming trend above the tropical mean.

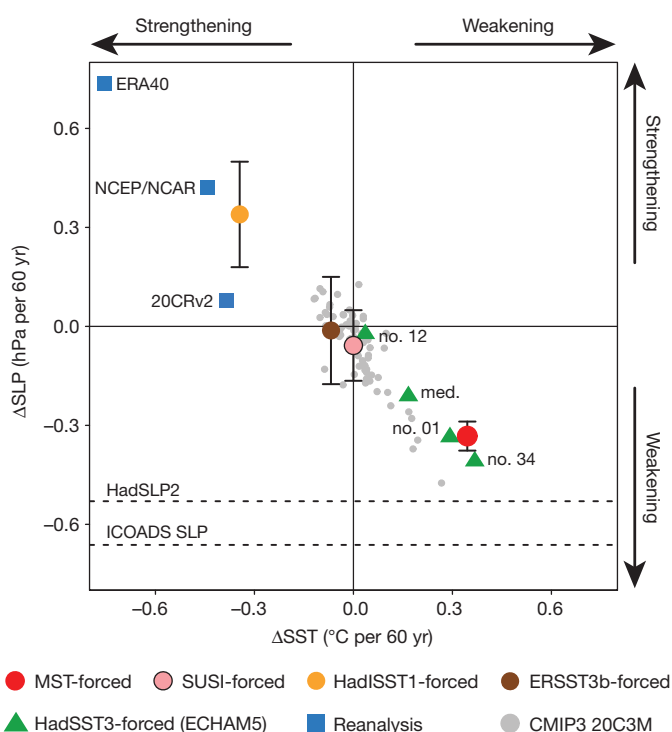


Figure 3 | Scatter diagram of ΔSST and ΔSLP changes in AGCM and CMIP3 model simulations and reanalysis data sets. Confidence intervals (error bars) for AGCM experiments are computed from the standard deviation of four ensemble members, at the two-sided $P = 0.01$ level. The twentieth-century trends are estimated from each ensemble member of the CMIP3 20C3M experiment, and scaled to 60-yr trends. Dashed lines indicate ΔSLP changes calculated from ICOADS and HadSLP2. A correlation coefficient between ΔSLP and ΔSST for AGCM experiments and reanalyses (CMIP3) is -0.97 (-0.83), significant at $P < 0.01$. See key at bottom for meaning of symbols.

Indo-Pacific, rather than to the muted hydrological cycle under uniform warming^{5,6}. This SST pattern mechanism itself is not new^{1,11,14,23} but its dominance in tropical climate change over the past six decades is surprising, given the prevailing assumption that the Walker circulation slowdown occurs without the zonal SST gradient change. Conversely, the reduced zonal SST gradient is the response to the weakened easterly trades⁸. The circular results suggest a positive feedback between the Walker circulation and SST gradient. In light of modest changes in the 20C3M simulations (Fig. 3), the observed changes discussed here probably include the influence of natural variability superimposed on externally forced response^{9,24}. The majority of CMIP phase 5 (CMIP5) models project a weakened SST gradient across the equatorial Indo-Pacific in response to increased greenhouse-gas forcing, but its magnitude is not significant in the twentieth century (Supplementary Fig. 5). On the other hand, the increased aerosols due to economic development and biomass burning may also have contributed to the reduced SST warming over the Indo-western Pacific region²⁵. An open question still remains: to what degree does anthropogenic forcing contribute to the observed changes? A conclusive attribution will have to wait for longer observations and the growth of the forced response.

Research into future climate change relies heavily on numerical models, but model skill in simulating long-term changes is largely untested. Validating climate models against observed change is critical in building confidence in their future projections. The slowdown of the Walker circulation is a robust change observed for the past century and serves a useful testbed for model evaluation. Our results show that AGCMs are successful in reproducing the Walker circulation slowdown, provided that the correct pattern of SST warming is prescribed. Taken together, our SST analysis and AGCM experiments indicate a weakening of zonal SST gradient across the equatorial Indo-Pacific over the past six decades. Our study highlights uncertainties in the tropical Indo-Pacific warming pattern and their effect on the Walker circulation change. The effect of SST warming uncertainty may extend remotely to densely populated regions such as East Asia through the atmospheric bridge²⁶ (Supplementary Fig. 4b). Our analysis suggests

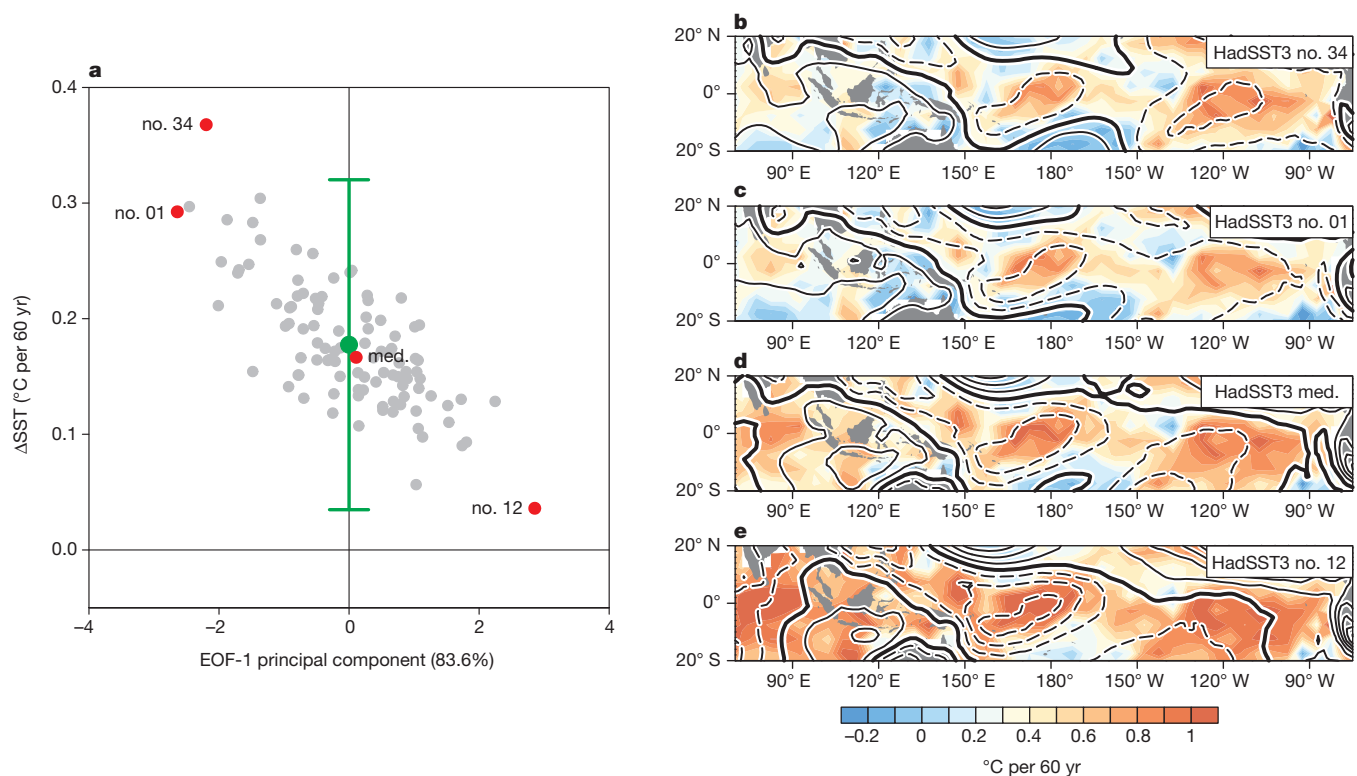


Figure 4 | Uncertainty in HadSST3 trends. **a**, Scatter diagram of ΔSST change and the principal component of the leading EOF for the tropical Indo-Pacific SST trend patterns. The median and extreme realizations no. 01, no. 12, no. 34 of HadSST3 are indicated in red circles. Green circle indicates the ensemble mean of the HadSST3 realizations with the confidence interval at the

that atmospheric reanalyses are also subject to errors of SST warming pattern. Their judicious use is advised in studying long-term changes. Reducing biases and uncertainty in regional patterns of SST warming is a key step for improving the reanalyses and interpretation of long-term climate change.

METHODS SUMMARY

The MST trends were calculated for 1950–2009 as an unweighted average of the bucket SST and NMAT trends from ICOADS release 2.5 (ref. 17) and the Met Office Historical Marine Air Temperature 43N (MOHMA4; ref. 15). We performed a set of two AGCM experiments, in which the 30-yr-scaled MST trend pattern was either added to or subtracted from the HadISST1 climatology over the tropics only (20°S–20°N). These two experiments were also forced with greenhouse-gas concentrations averaged in the 2000s and 1950s, respectively. By taking the difference between the two experiments, we extracted atmospheric changes over the past 60 yr. To investigate the effect of SST warming patterns, the same sets of AGCM experiments were forced by HadISST1, ERSST3b and HadSST3 trend patterns and SUSI ($\sim 0.5^{\circ}\text{C}$ increase per 60 yr). To obtain an unbiased estimate of the atmospheric response, we used four AGCMs: the US National Oceanic and Atmospheric Administration (NOAA) Geophysical Fluid Dynamics Laboratory (GFDL) AM2.1 (ref. 27), the Max Planck Institute (MPI) ECHAM5 (ref. 28), the NCAR Community Atmospheric Model (CAM) version 3 (ref. 29) and 4 (ref. 30). The AM2.1 (CAM4) uses a finite-volume grid of $2.5^{\circ} \times 2^{\circ}$ ($2.5^{\circ} \times 1.9^{\circ}$) longitude–latitude and 24 (26) vertical levels, and the ECHAM5 (CAM3) uses a T42 grid and 19 (26) vertical levels. For each experiment the models were integrated for 41 yr with the first year of integration discarded as a spin-up. Further details of observational data sets, AGCM experiments and analysis methods are discussed in Methods.

Full Methods and any associated references are available in the online version of the paper.

Received 21 April; accepted 11 September 2012.

- Deser, C., Phillips, A. S. & Alexander, M. A. Twentieth century tropical sea surface temperature trends revisited. *Geophys. Res. Lett.* **37**, L10701, <http://dx.doi.org/10.1029/2010GL043321> (2010).

two-sided $P = 0.01$ level (green error bar). **b–e**, SST (shading) and SLP (contour interval 0.15 hPa per 60 yr) changes from ECHAM5 experiments forced with HadSST3 realizations no. 34 (**b**), no. 01 (**c**), median (**d**) and no. 12 (**e**). Positive (negative) contours are solid (dashed) lines, and zero contours are thickened.

- Kennedy, J. J., Rayner, N. A., Smith, R. O., Parker, D. E. & Saunby, M. Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 2. Biases and homogenization. *J. Geophys. Res.* **116**, D14104, <http://dx.doi.org/10.1029/2010JD015220> (2011).
- Xie, S.-P. et al. Global warming pattern formation: sea surface temperature and rainfall. *J. Clim.* **23**, 966–986 (2010).
- Shin, S. I. & Sardeshmukh, P. D. Critical influence of the pattern of tropical ocean warming on remote climate trends. *Clim. Dyn.* **36**, 1577–1591 (2011).
- Vecchi, G. A. et al. Weakening of tropical Pacific atmospheric circulation due to anthropogenic forcing. *Nature* **441**, 73–76 (2006).
- Held, I. M. & Soden, B. J. Robust responses of the hydrological cycle to global warming. *J. Clim.* **19**, 5686–5699 (2006).
- Vecchi, G. A. & Soden, B. J. Global warming and the weakening of the tropical circulation. *J. Clim.* **20**, 4316–4340 (2007).
- Tokina, H. et al. Regional patterns of tropical Indo-Pacific climate change: evidence of the Walker circulation weakening. *J. Clim.* **25**, 1689–1710 (2012).
- Power, S. B. & Kociuba, G. What caused the observed twentieth-century weakening of the Walker circulation? *J. Clim.* **24**, 6501–6514 (2011).
- DiNezio, P., Clement, A. & Vecchi, G. A. Reconciling differing views of tropical Pacific climate change. *Eos* **91**, 141–142 (2010).
- Clement, A., DiNezio, P. & Deser, C. Rethinking the ocean's role in the Southern Oscillation. *J. Clim.* **24**, 4056–4072 (2011).
- Meng, Q. et al. Twentieth century Walker circulation change: data analysis and model experiments. *Clim. Dyn.* **38**, 1757–1773 (2012).
- Meeth, G. A. et al. in *Climate Change 2007: The Physical Science Basis* (eds Solomon, S. et al.), 747–845 (Cambridge Univ. Press, 2007).
- Bunge, L. & Clarke, A. J. A verified estimation of the El Niño index Niño-3.4 since 1877. *J. Clim.* **22**, 3979–3992 (2009).
- Rayner, N. A. et al. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.* **108**, 4407, <http://dx.doi.org/10.1029/2002JD002670> (2003).
- Smith, T. M., Reynolds, R. W., Peterson, T. C. & Lawrimore, J. Improvements to NOAA's historical merged land-ocean surface temperature analysis (1880–2006). *J. Clim.* **21**, 2283–2296 (2008).
- Woodruff, S. D. et al. ICOADS Release 2.5: extensions and enhancements to the surface marine meteorological archive. *Int. J. Climatol.* **31**, 951–967 (2011).
- Kalnay, E. et al. The NCEP/NCAR 40-year reanalysis project. *Bull. Am. Meteorol. Soc.* **77**, 437–471 (1996).
- Uppala, S. M. et al. The ERA-40 re-analysis. *Q. J. R. Meteorol. Soc.* **131**, 2961–3012 (2005).
- Compo, G. P. et al. The twentieth century reanalysis project. *Q. J. R. Meteorol. Soc.* **137**, 1–28 (2011).

21. Johnson, N. C. & Xie, S.-P. Changes in the sea surface temperature threshold for tropical convection. *Nature Geosci.* **3**, 842–845 (2010).
22. Kennedy, J. J., Rayner, N. A., Smith, R. O., Parker, D. E. & Saunby, M. Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 1. Measurement and sampling uncertainties. *J. Geophys. Res.* **116**, D14103, <http://dx.doi.org/10.1029/2010JD015218> (2011).
23. Deser, C. & Phillips, A. S. Simulation of the 1976/77 climate transition over the North Pacific: sensitivity to tropical forcing. *J. Clim.* **19**, 6170–6180 (2006).
24. Meehl, G. A., Hu, A. X. & Santer, B. D. The mid-1970s climate shift in the Pacific and the relative roles of forced versus inherent decadal variability. *J. Clim.* **22**, 780–792 (2009).
25. Ramanathan, V. *et al.* Atmospheric brown clouds: impacts on South Asian climate and hydrological cycle. *Proc. Natl Acad. Sci. USA* **102**, 5326–5333 (2005).
26. Xie, S.-P. *et al.* Decadal shift in El Niño influences on Indo-western Pacific and East Asian climate in the 1970s. *J. Clim.* **23**, 3352–3368 (2010).
27. The GFDL Global Atmospheric Model Development Team. The new GFDL global atmosphere and land model AM2–LM2: evaluation with prescribed SST simulations. *J. Clim.* **17**, 4641–4673 (2004).
28. Roeckner, E. *et al.* *The Atmospheric General Circulation Model ECHAM5. Part I: Model Description* (MPI report 349, Max Planck Institute, 2003); available at http://www.mpimet.mpg.de/fileadmin/publikationen/Reports/max_scirep_349.pdf (2003).
29. Collins, W. D. *et al.* The formulation and atmospheric simulation of the Community Atmosphere Model version 3 (CAM3). *J. Clim.* **19**, 2144–2161 (2006).
30. Neale, R. B. *et al.* The mean climate of the Community Atmosphere Model (CAM4) in forced SST and fully coupled experiments. *J. Clim.* (submitted).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank A. Timmermann for discussions, and X. T. Zheng and J. Ma for data processing. The work was supported by JAMSTEC, the National Basic Research Program of China (2012CB955600), NASA, NSF and NOAA. NCAR is sponsored by NSF.

Author Contributions H.T., Y.K. and Y.M.O. designed and performed model experiments. S.-P.X. and C.D. supervised the work. H.T. analysed observations and model simulations. H.T., S.-P.X., C.D. and Y.M.O. wrote the manuscript, with feedback from all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.T. (tokenaga@hawaii.edu) and S.-P.X. (xie@hawaii.edu).

METHODS

SLP, SST and NMAT. For SLP, we used the ICOADS release 2.5 (ref. 17) and the Hadley Centre's mean SLP data set version 2 (HadSLP2; ref. 31). For SST, we used HadISST1 (ref. 15), ERSST3b (ref. 16), HadSST3 (refs 2, 22) and bucket-sampled SSTs from ICOADS. For surface air temperature, we analysed only night-time data because daytime observations are strongly biased by heating effects over ship decks. For NMAT, ICOADS and the Met Office Historical Marine Air Temperature (MOHMAT) 43N (ref. 15) were used. All ICOADS data have been trimmed using the climatological 4.5 standard-deviation limit that identifies potential outliers and then gridded on a 4° grid with an unweighted box average⁸. Missing data in ICOADS were filled with linear interpolation across 'no data' grids not exceeding 5 points in longitude and 3 points in latitude. This interpolation was applied twice to obtain basin-scale characteristics. For the MST trend patterns, we first calculated trends in bucket SST and NMAT separately, and then constructed an unweighted average of their trend patterns. To improve the readability, the MST trend patterns were spatially smoothed with a 5×5 point Gaussian filter. ICOADS (<http://rda.ucar.edu/datasets/ds540.0/>), HadSLP2 (<http://www.metoffice.gov.uk/hadobs/hadslp2/>), HadISST1 (<http://www.metoffice.gov.uk/hadobs/hadisst1/>), ERSST3b (<http://www.ncdc.noaa.gov/ersst/>), HadSST3 (<http://www.metoffice.gov.uk/hadobs/hadst3/>), MOHMAT43N (<http://www.metoffice.gov.uk/hadobs/mohmat/>) are all available online.

Surface wind. The Wave- and Anemometer-based Sea-surface Wind (WASWind) data set³² (<http://iprc.soest.hawaii.edu/users/tokinaga/waswind.html>) was used. This data set substantially reduces spurious upward trends in the ICOADS raw winds by applying bias corrections with a stability-dependent height correction for measured winds and using wind-wave height. The obtained trend patterns are in good agreement with SLP measurements since the 1950s and satellite measurements for recent decades, illustrating the utility for climate change studies.

Precipitation and cloudiness. We used four rain-gauge based monthly mean gridded products available at NOAA³³ (<http://www.esrl.noaa.gov/psd/data/gridded/data.prcel.html>), the Global Precipitation Climatology Centre³⁴ (ftp://ftp.dwd.de/pub/data/gpcc/html/fulldata_v6_doi_download.html), the University of Delaware³⁵ (http://climate.geog.udel.edu/~climate/html_pages/Global2_Ts_2009/Global_p_ts_2009.html), and the University of East Anglia Climate Research Unit³⁶ (<http://www.cru.uea.ac.uk/cru/data/precip/>). All data were regridded onto a $3.75^\circ \times 2.5^\circ$ longitude–latitude grid, and precipitation trends were calculated for 1950–2007 from an unweighted average of the four products. Marine cloudiness trends were calculated for 1950–2009 based on ICOADS. To eliminate the bias in ship-observed marine cloudiness, we removed the tropical (30°N – 30°S) mean trend from each oceanic grid box¹.

Estimate of observed trends. All trends were calculated using only well-sampled grid boxes that contain monthly means for more than 75% of the total months. To suppress interannual variability for each month, all yearly data were smoothed with a five-point binominal temporal filter before the trend analysis. We estimated long-term trends and their statistical significance with the Sen median slope³⁷ and the Mann-Kendall³⁸ test, respectively, non-parametric methods less affected by outliers. The analysis period is basically from 1950 to 2009, but some data sets are not available for the entire period. In such cases, we scaled the trends to the 60-yr changes. Annual mean trends were presented in all figures.

AGCM experiments. We used the NOAA GFDL AM2.1 (ref. 27), the MPI ECHAM5 (ref. 28), the NCAR CAM3 (ref. 29) and CAM4 (ref. 30). The AM2.1

(CAM4) uses the finite-volume grid of $2.5^\circ \times 2^\circ$ ($2.5^\circ \times 1.9^\circ$) longitude–latitude and 24 (26) vertical levels, and the ECHAM5 (CAM3) uses a T42 grid and 19 (26) vertical levels. For each model and SST data set, we performed a set of two experiments, in which the 30-yr-scaled SST trend pattern was either added to or subtracted from the SST climatology over the tropics only (20°S – 20°N). Two SST forcing fields were derived as follows:

$$\text{SST}_{\text{warm}} = \overline{\text{SST}} + 30 \text{ SST}'$$

$$\text{SST}_{\text{cold}} = \overline{\text{SST}} - 30 \text{ SST}'$$

where $\overline{\text{SST}}$ is the monthly SST climatology obtained from HadISST1, and SST' is the monthly mean SST trends per year estimated from each SST data set. Both $\overline{\text{SST}}$ and SST' were calculated for 1950–2009. For the SUSI experiment, we used a tropical mean SST trend of 0.0083°C increase per year ($\sim 0.5^\circ\text{C}$ per 60 yr) based on the HadSST3 estimate. All SST forcing is yearly cycling data so that does not include interannual variability. The SST_{warm} and SST_{cold} experiments were also forced with greenhouse-gas concentrations (CO_2 , CH_4 , N_2O , CFC11, CFC12 and O_3) averaged in the 2000s and 1950s, respectively. By taking the difference between the two experiments, we extracted atmospheric changes over the past 60 yr. For each experiment the models were integrated for 41 yr with the first year of integration discarded as a spin-up. We used only ECHAM5 for HadSST3-forced experiments. The statistical significance for the difference between SST_{warm} and SST_{cold} experiments was estimated with the two-sided Student's *t*-test.

Other data sets. We used three atmospheric reanalysis products: the NCEP–NCAR reanalysis¹⁸ for 1950–2009 (<http://rda.ucar.edu/pub/reanalysis/index.html>), the ERA-40 reanalysis¹⁹ for 1958–2001 (<http://www.ecmwf.int/products/data/archive/descriptions/e4/index.html>), and 20CRv2 (ref. 20) for 1950–2008 (http://www.esrl.noaa.gov/psd/data/gridded/data.20thC_ReanV2.html). We also used the CMIP3 'Climate of the twentieth century experiments (20C3M)' (24 models) for 1900–1999, the CMIP5 historical run (22 models) for 1900–2005, and the CMIP5 high emission scenario (RCP8.5; 22 models) and medium mitigation scenario (RCP4.5; 22 models) for 2006–2098.

31. Allan, R. & Ansell, T. A new globally complete monthly historical gridded mean sea level pressure dataset (HadSLP2): 1850–2004. *J. Clim.* **19**, 5816–5842 (2006).
32. Tokinaga, H. & Xie, S.-P. Wave- and anemometer-based sea surface wind (WASWind) for climate change analysis. *J. Clim.* **24**, 267–285 (2011).
33. Chen, M. Y., Xie, P. P., Janowiak, J. E. & Arkin, P. A. Global land precipitation: a 50-yr monthly analysis based on gauge observations. *J. Hydrometeorol.* **3**, 249–266 (2002).
34. Schneider, U., Fuchs, T., Meyer-Christoffer, A. & Rudolf, B. *Global Precipitation Analysis Products of the GPCC* (Global Precipitation Climatology Centre, 2008); available at ftp://ftp-anon.dwd.de/pub/data/gpcc/PDF/GPCC_intro_products_2008.pdf (2008).
35. Willmott, C. J. & Robeson, S. M. Climatologically aided interpolation (CAI) of terrestrial air temperature. *Int. J. Climatol.* **15**, 221–229 (1995).
36. Hulme, M., Osborn, T. J. & Johns, T. C. Precipitation sensitivity to global warming: comparison of observations with HadCM2 simulations. *Geophys. Res. Lett.* **25**, 3379–3382 (1998).
37. Sen, P. K. Estimates of the regression coefficient based on Kendall's tau. *J. Am. Stat. Assoc.* **63**, 1379–1389 (1968).
38. Kendall, M. G. *Rank Correlation Methods* (Griffin, 1975).

The global diversity of birds in space and time

W. Jetz^{1*}, G. H. Thomas^{2*}, J. B. Joy^{3*}, K. Hartmann⁴ & A. O. Mooers³

Current global patterns of biodiversity result from processes that operate over both space and time and thus require an integrated macroecological and macroevolutionary perspective^{1–4}. Molecular time trees have advanced our understanding of the tempo and mode of diversification^{5–7} and have identified remarkable adaptive radiations across the tree of life^{8–10}. However, incomplete joint phylogenetic and geographic sampling has limited broad-scale inference. Thus, the relative prevalence of rapid radiations and the importance of their geographic settings in shaping global biodiversity patterns remain unclear. Here we present, analyse and map the first complete dated phylogeny of all 9,993 extant species of birds, a widely studied group showing many unique adaptations. We find that birds have undergone a strong increase in diversification rate from about 50 million years ago to the near present. This acceleration is due to a number of significant rate increases, both within songbirds and within other young and mostly temperate radiations including the waterfowl, gulls and woodpeckers. Importantly, species characterized with very high past diversification rates are interspersed throughout the avian tree and across geographic space. Geographically, the major differences in diversification rates are hemispheric rather than latitudinal, with bird assemblages in Asia, North America and southern South America containing a disproportionate number of species from recent rapid radiations. The contribution of rapidly radiating lineages to both temporal diversification dynamics and spatial distributions of species diversity illustrates the benefits of an inclusive geographical and taxonomical perspective. Overall, whereas constituent clades may exhibit slowdowns^{10,11}, the adaptive zone into which modern birds have diversified since the Cretaceous may still offer opportunities for diversification.

Birds (class Aves) constitute a fascinating and widely-studied radiation. Analyses based on a very incomplete ‘tapestry’ phylogeny¹² suggested higher speciation and diversification rates in the tropics and in South relative to North America^{13–15}. In addition, numerous geographically disparate clades are considered exceptional radiations, including both New and Old World warblers in the Northern Hemisphere^{10,16}, island radiations such as Darwin’s finches¹⁷, and the explosively diversifying white-eyes that span much of the southern Old World¹⁸. The prevalence and implications of such rapidly radiating clades have not been put in broader context: how characteristic are shifts in diversification, and to what extent do clade-specific and tree-wide variation in diversification rate contribute to diversity dynamics across the extant tree of the entire class? Where do these radiating lineages occur, and how much do they contribute to current-day patterns of diversity in the highly diverse tropics compared to relatively depauperate higher latitudes?

We address these questions using the first set of complete phylogenies of extant bird species (9,993 species, see Methods), compiled in a Bayesian framework, and a new species-level measure of past diversification rate. The phylogeny builds on previously established deeper relationships and combines molecular data for 6,663 species with taxonomic constraints for data-deficient species to more fully account

for phylogenetic uncertainty. Lineages-through-time⁸ and novel diversification-rates-through-time plots (Fig. 1) indicate that net diversification leading to extant lineages generally increased from approximately

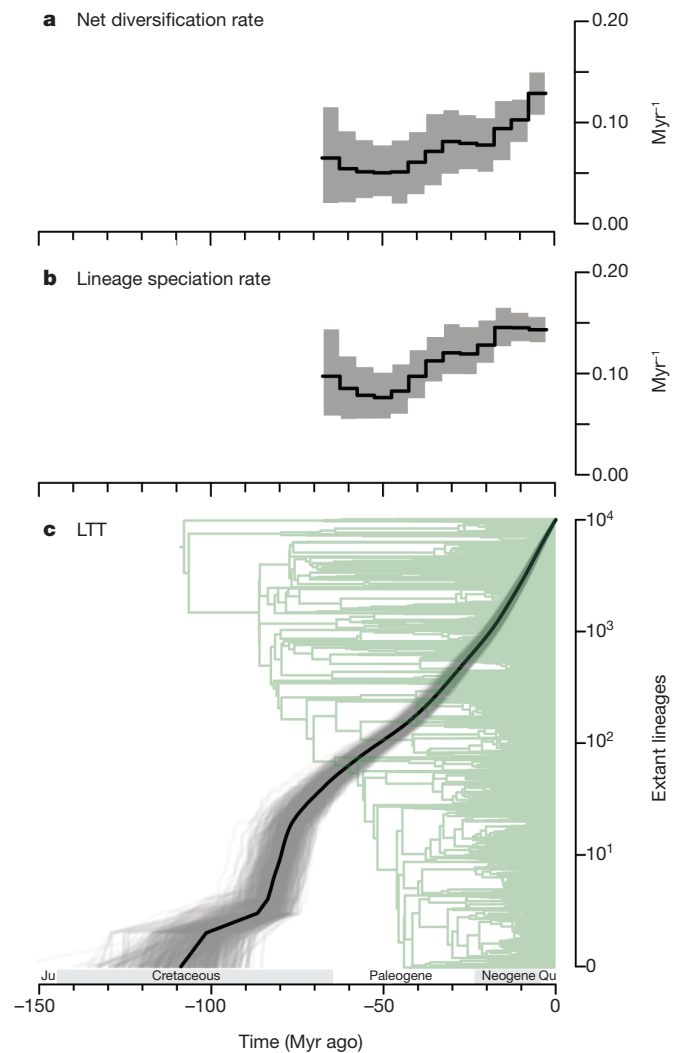


Figure 1 | Diversification of all birds through time. **a, b,** Estimates of the tree-wide lineage net diversification (speciation – extinction) rate (**a**) and speciation rate (**b**), calculated in 5 million year intervals (line segments). These are estimated to be very similar (see Supplementary Discussion). The shaded region represents the area between the 5th and 95th quantiles for 525 assessed trees with the mean rate traced in black. Intervals outside 67.5 and 2.5 Myr ago are not shown due to lack of data (≤ 30 lineages per interval) and the difficulty of accounting for ongoing speciation events, respectively. **c,** Lineage-through time plot for 1,000 trees (in grey), with mean waiting times to speciation in black. Green background is the tree depicted in Fig. 2. Geologic time periods are delineated at the bottom of the plot. Ju, Jurassic period; Qu, Quaternary period.

¹Department of Ecology and Evolutionary Biology, Yale University, 165 Prospect Street, New Haven, Connecticut 06520-8106, USA. ²Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK. ³Department of Biological Sciences, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia V5A 1S6, Canada. ⁴Institute for Marine and Antarctic Studies, University of Tasmania, Private Bag 49, Hobart, Tasmania 7001, Australia.

*These authors contributed equally to this work.

50 million years (Myr) ago to at least 5 Myr ago. This contrasts markedly with patterns of decreasing net diversification reported for smaller clades for which a model of niche-filling coupled with long-term lineage competition¹⁹ has been invoked^{9,10,20–22}.

By far the best model for diversification (median Akaike Information Criterion improvement over all other models $\gg 1,700$; Supplementary Discussion and Table 1) is one where a subset of individual clades (identified in Fig. 2 and Table 1) are assigned their own constant and elevated diversification rates. This confirms a space- and time-dependent history of avian diversification, with most of the consistently identified rate increases occurring within the last ~ 50 Myr. Several of these shifts confirm previously identified highly diverse clades with key morphological and behavioural innovations or environmental opportunities, including hummingbirds, parrots and a number of songbird lineages. Some additional young and rapidly radiating groups with rates of per-lineage species accumulation ≥ 0.25 species per million years (Table 1) stand out. These include the previously unrecognized rapid radiations of both ducks and geese, and of select gulls, as well as recognized rapid radiations of ovenbirds plus woodcreepers²³, and of white-eyes¹⁸, all with extant diversity ranging from 44 to 300 species. Intriguingly, these recently rapidly radiating clades show no obvious restriction to parts of the phylogenetic tree (Fig. 2) or landmasses. Rapidly and slowly radiating clades are highly interspersed. Collectively, these tree-wide distributed bursts produce the overall average increase in tree-wide diversification rate (Fig. 1a). Importantly, these novel patterns assume rate constancy within clades, and so integrate (rather than conflict with) changing rates associated with, for example, diversity- or time-dependent slowdowns at smaller scales within individual clades^{9,10,20}.

To understand the consequences of rate variation across the tree for the present-day geographic variation in avian diversity in more detail, we calculate a new metric of species-level lineage diversification rate (DR) for every species. The measure captures the longer-term splitting rate leading to a given species, made possible by the full resolution of the phylogeny. It is directly related to clade level diversification rate measures (see Supplementary Methods), but offers species-level resolution. Mean species values range from 0.01 to 4.66 species Myr⁻¹ and roughly follow a lognormal distribution (Fig. 2, centre) with a geometric mean of 0.157 species Myr⁻¹ and only limited variation owing to

captured phylogenetic uncertainty (Supplementary Discussion Fig. 4). The species level detail illustrates the marked heterogeneity in diversification rates beyond the main shifts and identifies numerous 'hot' sections of recent rapid radiations dispersed widely across the avian tree of life.

This heterogeneity extends to the geographical prevalence of species with a signature of high past diversification. We find that the main geographic differences in diversification rate are east–west hemispheric, rather than latitudinal^{14,15} (see below). Average rates are distinctly lower in the Eastern (east of 28° W, DR = 0.145, $N = 5,810$) compared to the Western Hemisphere (DR = 0.177, $N = 4,183$, $P_{\text{avg}} < 0.05$, $f(P < 0.05) = 73/100$, Fig. 3). In contrast, there are no significant difference in rates ($P = 0.69$, $f(P < 0.05) = 0/100$) between Northern (DR = 0.161, $N = 5,084$) and Southern Hemisphere species (DR = 0.154, $N = 4,909$). Intriguingly, avian assemblages in Australia, Southeast Asia, Africa and Madagascar are characterized by particularly low average rates (below approximately 0.12 species Myr⁻¹ in most locations, compared to a global mean of 0.16). These regions also harbour substantially fewer than expected of the 25% of species with highest diversification rate ($< 15\%$ in most assemblages, Fig. 3d). One hypothesis for the patterns in Australia and Africa is regional density-dependent diversification, with early filling of ecological space by ancient radiations in their regions of origin²⁴. Generally, high diversification rates and large relative prevalence of top diversification rate species are found throughout higher-latitude North America, parts of north Asia and southwest South America—the main breeding areas of several of the rapidly radiating clades identified in Table 1 (including warblers, ducks, gulls and woodpeckers). These regions have all been characterized by strong climatic fluctuations from the Pliocene to the present (with, for example, notable emergence and size fluctuations of wetland and forest breeding habitats), which supports the suggestion that the geography of past climate dynamics has had a major role for today's diversity patterns¹⁴. Isolated locations such as islands have been proposed to strongly facilitate rapid radiations through the reduction in gene flow and novel ecological opportunities they often present after initial colonization^{17,25}. We confirm this globally, as non-pelagic birds with more than half of their range on islands ($N = 1,085$, DR = 0.203) have much larger diversification rates than predominantly mainland birds ($N = 8,629$, DR = 0.153; $P_{\text{avg}} < 0.001$,

Table 1 | Rate shifts found in at least 25% of sampled trees.

Node	English	Scientific	Age	r	Species
G	Select gulls	Select Laridae	4.6 (0.0)	0.74	44
U	White-eyes, select babblers	Zosteropidae, select Timaliidae	8.1 (0.1)	0.49	113
N	Ovenbirds, woodcreepers	Dendrocolaptidae, Furnariidae	17.8 (0.1)	0.28	285
[O]	- Ovenbirds	- Furnariidae	13.6 (0.1)	0.31	223
C	Select ducks, geese	Anseriformes	10.8 (0.1)	0.28	109
T	Babblers, white-eyes	Timaliidae, Zosteropidae	17.7 (0.1)	0.27	355
Y	Weavers, estrildid finches	Ploceidae, Estrildidae	21.9 (0.2)	0.24	278
H	Woodpeckers and allies	Ramphastidae, Indicatoridae, Picidae	36.4 (0.3)	0.17	361
[I]	- Woodpeckers	- Picidae	23.8 (0.2)	0.21	220
X	Passeroidea songbirds	Passeroidea	35.7 (0.1)	0.20	1,418
D	Select hummingbirds	Select Trochilidae	20.7 (0.1)	0.19	266
R	Select sylvioid songbirds	Select Sylvioidae	33.3 (0.1)	0.17	1,031
[S]	- Subset	- Subset	29.1 (0.1)	0.19	704
Q	Passerida songbirds	Passerida	44.7 (0.1)	0.16	3,574
M	New World suboscines	Tyrannides	52.4 (0.2)	0.16	1,227
P	Select Corvida songbirds	Select Corvida	32.1 (0.2)	0.16	442
F	Gulls, terns, auks, skuas, jaegers	Stercorariidae, Alcidae, Laridae	21.1 (0.1)	0.15	131
V	Muscipapoidea songbirds	Muscipapoidea	38.8 (0.1)	0.14	732
[W]	- subset	- Subset	29.4 (0.2)	0.15	603
J	Parrots, songbirds	Psittaciformes, Passeriformes	77.2 (0.2)	0.14	6,320
[L]	- Most songbirds	- Passeriformes	66.8 (0.2)	0.14	5,964
A	Fowl	Galloanserae	78.7 (0.2)	0.08	448
[B]	- Landfowl	- Galliformes	53.9 (0.3)	0.13	267
K	True parrots	Psittacidae	35.6 (0.3)	0.13	330
E	Pigeons, doves	Columbidae	33.4 (0.1)	0.12	289

Rate shifts were identified using MEDUSA⁵ (see Fig. 2 for position and prevalence). Pairs of nested shifts are combined when younger shifts (in square brackets) were depressed by the presence of older to $< 10\%$ occurrence. 'Node' refers to position on example tree (Fig. 2). 'Age' of node is the mean age in units million years (with s.e. calculated across trees) calculated across the posterior sample of trees and 'r' is the mean per lineage diversification rate (in units of Myr⁻¹, s.e. across trees < 0.01 for all) estimated for focal nodes of given species richness ('Species'). All clade-level rate estimates were inferred using Laser²⁸. The background rate as estimated with MEDUSA was $r = 0.056$ Myr⁻¹. Support for nodes was 1.00 for all except U (0.02), T (0.73), S (0.48), P (0.85) and E (0.93).

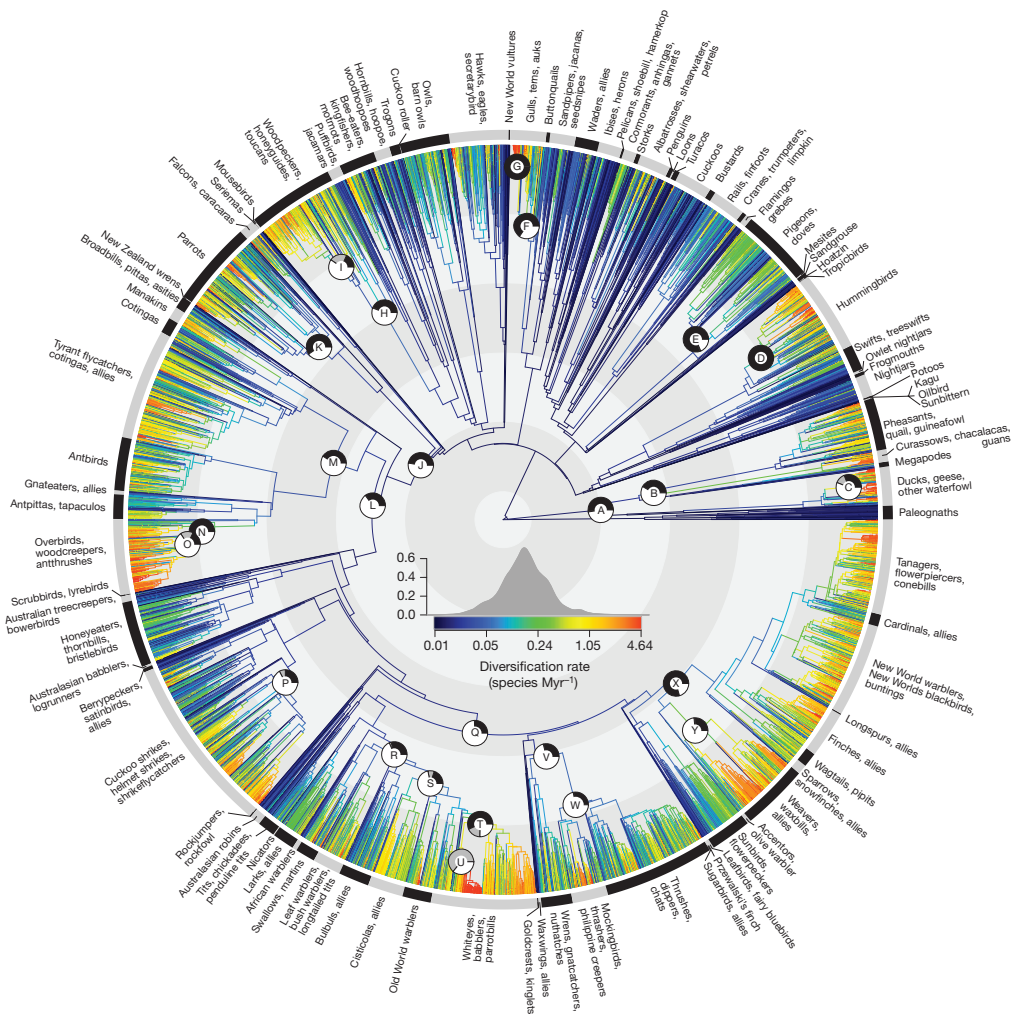


Figure 2 | Diversification across the avian tree. Diversification rate shifts identified by MEDUSA and the species level diversification rate metric (DR) are displayed on a representative avian tree. Nodes with shifts in diversification rate identified in at least 25% of the tested trees are indicated by pie-charts (labelled A–Y, see Table 1). Black and grey areas show the proportion of trees with a shift at the focal node and with shifts that are nested within (more recent than) the focal node, respectively. Shifts are counted only once (for example, shifts at node U do not contribute to the prevalence of nested shifts at node T). Prevalence of shifts may be lowered by a ‘trickle-down’ effect where combinations of nested shifts are rarely identified in the same tree (see also

Supplementary Discussion Fig. 6 for discussion). Branches are coloured according to the mean diversification rate of descendant branches. We colour branches for visualization purposes to highlight tree-wide variation in diversification rate and do not analyse values for internal branches. Diversification rate quantifies the splitting rate along branches leading to a species and offers species-level detail for clade-level diversification rate (see Supplementary Methods). The inset shows the scale and frequency distribution of diversification rate values across species. Concentric grey circles show time from the present in 20 million year intervals.

$f(P < 0.05) = 100/100$). Many islands, particularly those inhabited by passerines in the Pacific, North Atlantic and Caribbean, stand out for their high diversification rates, for example, as shown by the dramatic white-eye radiation (Table 1). As might be expected on the basis of species richness alone, passerines ($N = 5,966$, $DR = 0.179$) have on average much higher diversification rate values than non-passerines ($N = 4,027$, $DR = 0.131$), a difference captured by the diversification rate increase observed at the base of the Passeriformes (node ‘L’ in Table 1, Fig. 2). Accordingly, passerines have a strong influence on the geographic patterns of highest diversification rate (Fig. 3b, c). However, both major groups harbour high and low diversification rate species and both have major hemispheric differences as well as higher rates on islands. Intriguingly though, non-passerines, especially through groups such as water birds, and some gulls and woodpeckers, drive the higher rates in Asia and also show high rates in North America. Passerines show significant recent radiations in temperate Asia and North America. In particular due to the ovenbird and woodcreeper clades, passerines contribute strongly to high average diversification rates and present richness in South America, especially south

of the core rainforest belt. We propose that over the past 10–20 Myr the expansion of core temperate habitats⁴ as well as mountain uplifts and, more recently, the retreat of high-latitude glaciations (followed by range expansions and speciation), all contributed to these distinctive patterns.

These geographic patterns add an important dimension to perspectives that have emphasized latitude¹⁵ as a predictor of both diversification rate and standing diversity. In a highly simplifying global latitudinal view (Fig. 4), high diversification rate species in all tropical lowland regions seem generally outnumbered by species of lower diversification rate. Overall, the simple hypothesis that latitudinal diversity differences result from raw variation in net rates by latitude is rejected: mean diversification rate shows no significant trend with absolute latitude or between the tropics and higher latitudes (Fig. 4). This finding is consistent with the hypothesis that the greater combined age and expanse of tropical moist forests has facilitated greater species accumulation without necessarily facilitating greater rates of diversification⁴. The equatorial dip in average diversification rates is followed by a peak in the slightly drier and (over the past 10 Myr)

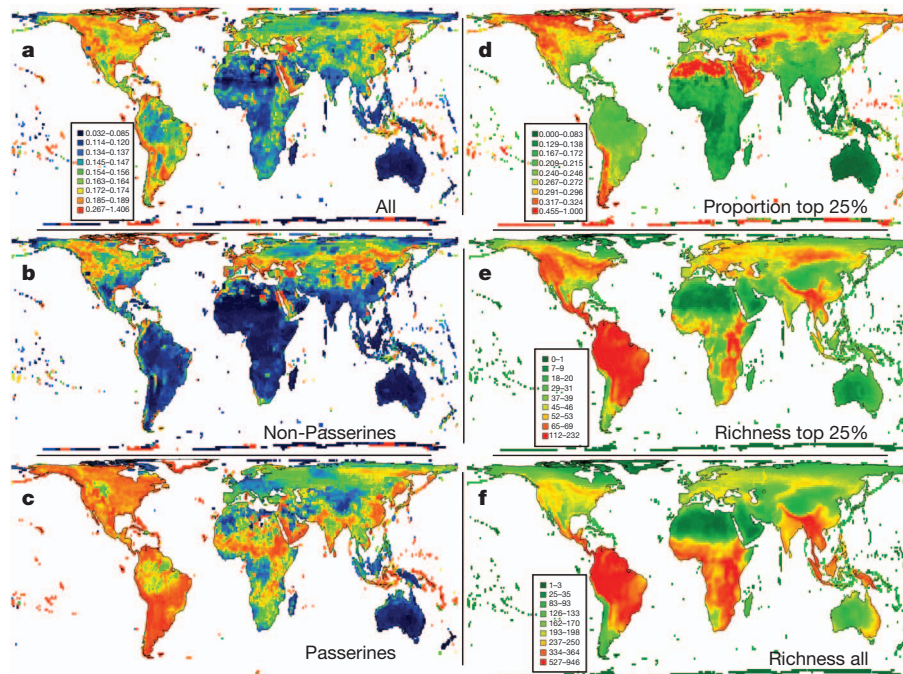


Figure 3 | Geographic variation in species-level lineage diversification rate and the richness of high-diversification rate species. a–c, Mean assemblage diversification rate (see Fig. 2), calculated as the geometric mean of all species in a grid cell assemblage, weighted by the inverse of their range size. a, All species; b, non-passerines; c, passerines. This visualization limits the overbearing

(pseudo-replicating) effect wide-ranging species have on perceived spatial patterns of assemblage summaries⁴. d–f, Relative (d) and absolute (e) richness of top 25% diversification rate species ($DR \geq 0.243$ species Myr^{-1}); f shows the richness of all 9,993 bird species for comparison. Grid cell size is 110×110 km for all panels (Behrman projection).

climatically more dynamic regions at around $10\text{--}15^\circ$ in both hemispheres and an overall decrease towards the high latitudes. However, as our spatial results indicate, the outcome of any such latitudinal comparisons, excepting perhaps a narrow equatorial band, will depend on hemisphere and region, and the inclusion, prevalence and current evolutionary dynamics of particular clades (see also ref. 3).

Our choice of a constant-rate birth model as a prior on diversification is conservative with respect to clade-level rate heterogeneity, increasing diversification towards the present and major hemispheric differences in lineage-diversification rate. Our results are also robust to differing phylogenetic hypotheses, including currently existing alternative backbone topologies and the inclusion of data-deficient species (see Supplementary Discussion). Indeed, our tree distribution was designed to both integrate the latest taxonomic data and to capture remaining phylogenetic uncertainty given current knowledge; this means it will not be suitable for some questions in avian systematics. In addition, the age of origination of crown group birds and the absolute diversification rates will be contingent on the specific fossil constraints used. Ongoing discovery of cryptic species may further affect rate estimates, especially in the tropics where taxonomists have been less active. Further, and importantly, our inference is necessarily limited to lineages leading to present-day species and cannot account for the non-random extinction of entire clades (for example, of the entire moa clade in New Zealand). Finally, the illustrated spatial patterns tie diversification rate to the occurrence of species today rather than the time of their origination when the distribution of bioclimatic zones and species may have been different. Nonetheless, the uncovered geographic heterogeneity in diversification rates seems pervasive and provides a first global integration of species-level diversification rate variation across both time and space. It may be that birds are distinctive in their pattern of an increasing diversification rate from ~ 50 Myr ago to the recent: the overall adaptive zone into which modern birds have diversified since at least the early Eocene may not be saturated, and opportunities for diversification may be expanding. Alternatively, if the evidence for saturation in other taxa is not due to sampling

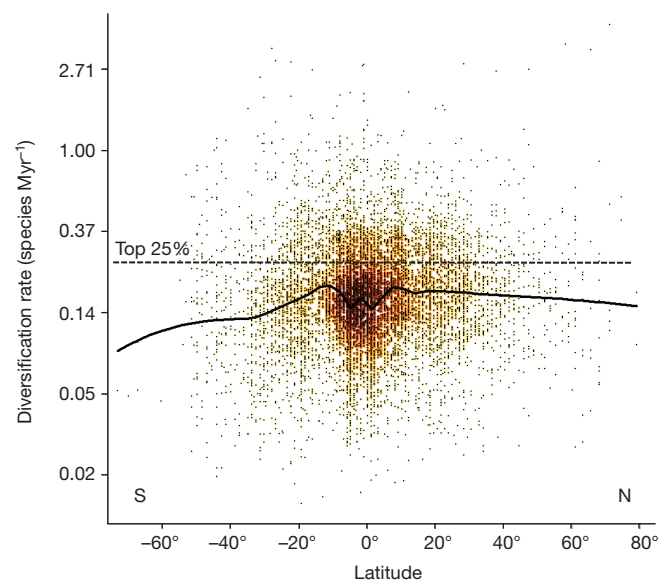


Figure 4 | Latitudinal gradient in species-level lineage diversification rate. Each black point represents a single species diversification rate (DR) at the centroid latitude of its global breeding range. Only the 75% species with small to moderate latitudinal extent ($<26.4^\circ$, $N = 7,493$) are included to ensure comparable centroid positions (see Supplementary Discussion Fig. 7 for very similar patterns and results including all species). There is no significant association between diversification rate and absolute centroid latitude ($P_{\text{avg}} = 0.51$, $f(P < 0.05) = 0/100$) or for intra- ($<23^\circ$ latitude) against extra-tropical centroid location ($P_{\text{avg}} = 0.16$, $f(P < 0.05) = 31/100$). The solid line is a loess smooth over all data (span = 0.2, degree = 2). The dashed line indicates the threshold identifying the quartile of species with highest diversification rate ($DR \geq 0.243$ Myr^{-1}). Darker brown shading highlights greater density of species points.

issues²⁶, then the difference may be due to the novel combined scale and resolution of our analyses, and smaller clades that are geographically or ecologically bounded may indeed saturate^{9–11}. We will need to compare global complete trees such as the one we present here with well-characterized component clades. We predict that constraint-induced slowdowns in net diversification will appear at smaller ecological and geographic scales, but that these do not combine to produce the same pattern at broader scales. This would be consistent with a classical view of multiple and ongoing, individually constrained adaptive radiations as a main driver of biodiversity¹⁹.

METHODS SUMMARY

Phylogenetic tree construction. Pseudo-posterior samples of complete avian trees were assembled as follows. (1) Every bird species was assigned to one of 158 clades identified using a backbone phylogeny²⁷. (2) Relaxed-clock trees were generated for each clade from sequence data. (3) Relaxed-clock trees for entire clades were generated combining species with and without genetic data: species without genetic information (3,330) were placed within their clade using constraint structures consistent with consensus trees from step (2) plus taxonomic information and branching times sampled from a pure birth model of diversification. (4) Final trees were assembled from the clade distributions plus samples of dated backbone trees from (one of two) distributions constructed using relaxed molecular clock techniques, 15 genes, ten fossil constraints and extensive topology constraints derived from published sources. For further information see Supplementary Methods, for trees, see Supplementary Results or <http://birdtree.org>.

Diversification analyses. We compared diversification models on 525 sampled trees: seven models fit smooth changes in rate²⁸, one searches for tree-wide rate shifts²⁹ and one identifies clades that have significantly elevated or depressed diversification rates relative to the entire tree⁵. To visualize diversification through time we estimated tree-wide speciation and extinction rates in five-million-year intervals using TreePar²⁹.

We measured the species-level lineage diversification rate (DR) for every species i as the inverse of its mean equal splits measure³⁰:

$$DR_i = \left(\sum_{j=1}^{N_i} l_j \frac{1}{2^{j-1}} \right)^{-1}$$

where N_i = number of edges on path from species i to the root and l_j = length of the edge j .

Comparative diversification rate analyses. For comparisons of diversification rate we used phylogenetic generalized least squares across 100 trees and list average P values and frequency (f) of trees with $P < 0.05$. For the island–mainland comparison we excluded 279 predominantly pelagic species.

Spatial analysis. We compiled non-invasive breeding distributions from the literature and select updates (see Supplementary Methods). We extracted these ranges over a 110 km equal area grid in Behrman projection excluding cells with $<30\%$ dry land or without off-shore islands, resulting in 2.43×10^6 occurrence records over 12,850 cells.

Received 5 April; accepted 28 September 2012.

Published online 31 October 2012.

- Mittelbach, G. G. *et al.* Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. *Ecol. Lett.* **10**, 315–331 (2007).
- Ricklefs, R. E. A comprehensive framework for global patterns in biodiversity. *Ecol. Lett.* **7**, 1–15 (2004).
- Linder, H. P. Plant species radiations: where, when, why? *Phil. Trans. R. Soc. B* **363**, 3097–3105 (2008).
- Jetz, W. & Fine, P. V. A. Global gradients in vertebrate diversity predicted by historical area-productivity dynamics and contemporary environment. *PLoS Biol.* **10**, e1001292 (2012).
- Alfaro, M. E. *et al.* Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *Proc. Natl Acad. Sci. USA* **106**, 13410–13414 (2009).
- Smith, S. A., Beaulieu, J. M., Stamatakis, A. & Donoghue, M. J. Understanding angiosperm diversification using small and large phylogenetic trees. *Am. J. Bot.* **98**, 404–414 (2011).
- Roelants, K. *et al.* Global patterns of diversification in the history of modern amphibians. *Proc. Natl Acad. Sci. USA* **104**, 887–892 (2007).
- Nee, S., Mooers, A. O. & Harvey, P. H. Tempo and mode of evolution revealed from molecular phylogenies. *Proc. Natl Acad. Sci. USA* **89**, 8322–8326 (1992).

- Phillimore, A. B. & Price, T. D. Density-dependent cladogenesis in birds. *PLoS Biol.* **6**, e71 (2008).
- Rabosky, D. L. & Lovette, I. J. Density-dependent diversification in North American wood warblers. *Proc. R. Soc. B* **275**, 2363–2371 (2008).
- Weir, J. T. Divergent timing and patterns of species accumulation in lowland and highland neotropical birds. *Evolution* **60**, 842–855 (2006).
- Sibley, C. G. & Ahlquist, J. E. *Phylogeny and Classification of Birds: a Study in Molecular Evolution* (Yale Univ. Press, 1990).
- Ricklefs, R. E. Global variation in the diversification rate of passerine birds. *Ecology* **87**, 2468–2478 (2006).
- Cardillo, M., Orme, C. D. L. & Owens, I. P. F. Testing for latitudinal bias in diversification rates: an example using New World birds. *Ecology* **86**, 2278–2287 (2005).
- Weir, J. T. & Schluter, D. The latitudinal gradient in recent speciation and extinction rates of birds and mammals. *Science* **315**, 1574–1576 (2007).
- Price, T., Lovette, I. J., Bermingham, E., Gibbs, H. L. & Richman, A. D. The imprint of history on communities of North American and Asian warblers. *Am. Nat.* **156**, 354–367 (2000).
- Grant, P. R. & Grant, B. R. *How and Why Species Multiply: the Radiation of Darwin's Finches* (Princeton Univ. Press, 2011).
- Moyle, R. G., Filardi, C. E., Smith, C. E. & Diamond, J. Explosive Pleistocene diversification and hemispheric expansion of a “great speciator”. *Proc. Natl Acad. Sci. USA* **106**, 1863–1868 (2009).
- Simpson, G. G. *The Major Features of Evolution* (Columbia Univ. Press, 1953).
- Etienne, R. S. *et al.* Diversity-dependence brings molecular phylogenies closer to agreement with the fossil record. *Proc. R. Soc. B* **279**, 1300–1309 (2012).
- Morlon, H., Parsons, T. L. & Plotkin, J. B. Reconciling molecular phylogenies with the fossil record. *Proc. Natl Acad. Sci. USA* **108**, 16327–16332 (2011).
- Ezard, T. H. G., Aze, T., Pearson, P. N. & Purvis, A. Interplay between changing climate and species' ecology drives macroevolutionary dynamics. *Science* **332**, 349–351 (2011).
- Derryberry, E. P. *et al.* Large-scale continental radiation: the neotropical ovenbirds and woodcreepers (Aves: Furnariidae). *Evolution* **65**, 2973–2986 (2011).
- Ericson, P. G. P. Evolution of terrestrial birds in three continents: biogeography and parallel radiations. *J. Biogeogr.* **39**, 813–824 (2012).
- Price, T. *Speciation in Birds* (Roberts, 2008).
- Cusimano, N. & Renner, S. S. Slowdowns in diversification rates from real phylogenies may not be real. *Syst. Biol.* **59**, 458–464 (2010).
- Hackett, S. J. *et al.* A phylogenomic study of birds reveals their evolutionary history. *Science* **320**, 1763–1768 (2008).
- Rabosky, D. L. LASER: a maximum likelihood toolkit for detecting temporal shifts in diversification rates from molecular phylogenies. *Evol. Bioinform. Online* **2**, 247–250 (2006).
- Stadler, T. Mammalian phylogeny reveals recent diversification rate shifts. *Proc. Natl Acad. Sci. USA* **108**, 6187–6192 (2011).
- Redding, D. W. & Mooers, A. O. Incorporating evolutionary measures into conservation prioritization. *Conserv. Biol.* **20**, 1670–1678 (2006).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank D. Redding for critical input in the early stages of this project; A. Mimoto, F. Ronqvist and M. Teslenko for help modifying MrBayes; I. Martyn for coding; R. Bowie, J. McGuire, A. Cooper, K. Burns and M. Sorenson among others, for unpublished phylogenetic material or information; M. Benton, T. Ezard, T. Price, M. Donoghue, J. Beaulieu, J. Belmaker, P. M. Hull, D. Field, N. Longrich, V. Saranathan, M. Steel, H. Morlon, J. Brown, A. Phillimore, R. Fitzjohn, R. Etienne, W. Stein and especially T. Stadler for data, important input and/or discussion; G. Smith, C. Schank, D. Thiele, T. M. Lee, F. La Sorte, C. Edwards, K. Ashton and J. Hazelhurst for help with spatial and phylogenetic data collection and management; C. Schank for help preparing the tree visualizations. This work was carried out using the BlueFern Supercomputing Facilities (<http://www.bluefern.canterbury.ac.nz>), University of Canterbury, the Advanced Computing Research Centre, University of Bristol (<http://www.bris.ac.uk/acrc/>) and the Interdisciplinary Research in Mathematics and Computer Sciences Centre, Simon Fraser University (<http://www.irmacs.sfu.ca>). This work was partly supported by NSF grants DBI 0960550 and DEB 1026764 and NASA Biodiversity Grant NNX11AP72G (W.J.); the Natural Environment Research Council (Postdoctoral Fellowship grant number NE/G012938/1 and the NERC Centre for Population Biology) (G.H.T.); and NSERC Canada, the Wissenschaftskolleg zu Berlin, the Yale Institute for Biospheric Sciences and Simon Fraser University (A.O.M.). Most importantly, we thank the many avian systematists and phylogeneticists who have contributed their data to public databases and so made our study possible.

Author Contributions W.J., A.O.M., and G.H.T. conceived of the study; K.H., W.J., J.B.J., A.O.M. and G.H.T. developed the methods; W.J., J.B.J. and G.H.T. collected the data; W.J., J.B.J. and G.H.T. conducted the analyses; W.J., J.B.J., A.O.M. and G.H.T. wrote the paper. W.J., J.B.J., G.H.T. and A.O.M. contributed equally to the study.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to W.J. (walter.jetz@yale.edu) or A.O.M. (amooers@sfu.ca).

An ultraviolet-radiation-independent pathway to melanoma carcinogenesis in the red hair/fair skin background

Devarati Mitra¹, Xi Luo², Ann Morgan¹, Jin Wang³, Mai P. Hoang⁴, Jennifer Lo¹, Candace R. Guerrero³, Jochen K. Lennerz⁵, Martin C. Mihm⁴, Jennifer A. Wargo⁶, Kathleen C. Robinson¹, Suprabha P. Devi¹, Jillian C. Vanover⁷, John A. D'Orazio⁷, Martin McMahon⁸, Marcus W. Bosenberg⁹, Kevin M. Haigis², Daniel A. Haber², Yinsheng Wang³ & David E. Fisher¹

People with pale skin, red hair, freckles and an inability to tan—the 'red hair/fair skin' phenotype—are at highest risk of developing melanoma, compared to all other pigmentation types¹. Genetically, this phenotype is frequently the product of inactivating polymorphisms in the melanocortin 1 receptor (*MC1R*) gene. *MC1R* encodes a cyclic AMP-stimulating G-protein-coupled receptor that controls pigment production. Minimal receptor activity, as in red hair/fair skin polymorphisms, produces the red/yellow pheomelanin pigment, whereas increasing *MC1R* activity stimulates the production of black/brown eumelanin². Pheomelanin has weak shielding capacity against ultraviolet radiation relative to eumelanin, and has been shown to amplify ultraviolet-A-induced reactive oxygen species^{3–5}. Several observations, however, complicate the assumption that melanoma risk is completely ultraviolet-radiation-dependent. For example, unlike non-melanoma skin cancers, melanoma is not restricted to sun-exposed skin and ultraviolet radiation signature mutations are infrequently oncogenic drivers⁶. Although linkage of melanoma risk to ultraviolet radiation exposure is beyond doubt, ultraviolet-radiation-independent events are likely to have a significant role^{1,7}. Here we introduce a conditional, melanocyte-targeted allele of the most common melanoma oncoprotein, BRAF^{V600E}, into mice carrying an inactivating mutation in the *Mclr* gene (these mice have a phenotype analogous to red hair/fair skin humans). We observed a high incidence of invasive melanomas without providing additional gene aberrations or ultraviolet radiation exposure. To investigate the mechanism of ultraviolet-radiation-independent carcinogenesis, we introduced an albino allele, which ablates all pigment production on the *Mclr*^{e/e} background. Selective absence of pheomelanin synthesis was protective against melanoma development. In addition, normal *Mclr*^{e/e} mouse skin was found to have significantly greater oxidative DNA and lipid damage than albino-*Mclr*^{e/e} mouse skin. These data suggest that the pheomelanin pigment pathway produces ultraviolet-radiation-independent carcinogenic contributions to melanomagenesis by a mechanism of oxidative damage. Although protection from ultraviolet radiation remains important, additional strategies may be required for optimal melanoma prevention.

To study the role of pigmentation in BRAF^{V600E} melanoma development, we used a series of genetically matched mice on the C57BL/6 background with various pigmentation phenotypes (Fig. 1a). To mimic dark-skinned individuals with a high eumelanin-to-pheomelanin ratio, we used mice with the wild-type C57BL/6 pigmentation phenotype ('black'). To mimic individuals with the red hair/fair skin phenotype who carry a high pheomelanin-to-eumelanin ratio, we used mice

with premature termination of the *Mclr* transcript (*Mclr*^{e/e}, 'red')⁸. To mimic individuals with albinism who have no melanin, we used mice with an inactivating mutation at the tyrosinase locus (*Tyr*^{c/c}, 'albino')⁹. Because tyrosinase is the initial and rate-limiting enzyme in melanin synthesis, albino melanocytes do not produce any pigment, but are normal in number and viability¹⁰.

We generated two variants of each pigmentation phenotype. One variant contains melanocytes in the dermis. A second matched variant contains transgenic stem cell factor expressed under the keratin 14 promoter (*K14-SCF*), which mimics SCF expression in human epidermal keratinocytes and results in epidermal melanocyte localization¹¹.

To create a genetic context primed for the induction of melanoma we also introduced into each of our six variants a system for inducible, melanocyte-specific expression of oncogenic BRAF^{V600E} (ref. 12). In humans, mice and zebrafish, expression of BRAF^{V600E} in melanocytes primarily causes benign nevi, rather than melanoma^{12–15}. In this context, malignant melanoma progression is thought to be constrained by

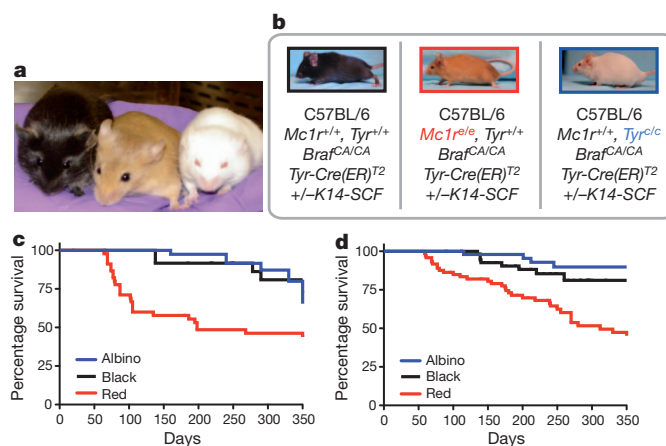


Figure 1 | Without ultraviolet radiation, *Braf*^{CA} red mice have an increased rate of melanoma development relative to black and albino *Braf*^{CA} animals. **a**, C57BL/6 pigmentation variants with epidermal melanocytes (*K14-SCF*). From left to right: black (wild type), red (*Mclr*^{e/e}) and albino (*Tyr*^{c/c}). **b**, Genotype of animals used for experimental studies. **c**, Percentage survival of pigmentation variants not carrying the *K14-SCF* transgene, that is, no epidermal melanocytes ($n_{\text{black}} = 28$, $n_{\text{red}} = 40$, $n_{\text{albino}} = 48$). $P_{\text{black-albino}} = 0.250$, $P_{\text{black-red}} = 0.003$, $P_{\text{albino-red}} = 0.003$. **d**, Percentage survival of pigmentation variants carrying the *K14-SCF* transgene, that is, epidermal melanocytes ($n_{\text{black}} = 49$, $n_{\text{red}} = 77$, $n_{\text{albino}} = 41$). $P_{\text{black-albino}} = 0.103$, $P_{\text{black-red}} = 0.009$, $P_{\text{albino-red}} < 0.0001$.

¹Cutaneous Biology Research Center, Massachusetts General Hospital, Charlestown, Massachusetts 02129, USA. ²Massachusetts General Hospital Cancer Center, Charlestown, Massachusetts 02129, USA. ³Department of Chemistry, University of California, Riverside, California 92521, USA. ⁴Department of Pathology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. ⁵Institute of Pathology, University Ulm, Ulm 89070, Germany. ⁶Department of Surgery, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. ⁷Markey Cancer Center, University of Kentucky School of Medicine, Lexington, Kentucky 40536, USA. ⁸Helen Diller Family Comprehensive Cancer Center, University of California San Francisco, San Francisco, California 94143, USA. ⁹Department of Dermatology, Yale University School of Medicine, New Haven, Connecticut 06520, USA.

oncogene-induced senescence¹⁵. Consistent with this, expression of BRAF^{V600E} in conjunction with silencing of *PTEN*, *TP53* or *CDKN2A* leads to development of malignant melanoma^{12–14}. However, spontaneous progression of BRAF^{V600E}-expressing melanocytes to malignant melanoma has been reported following a long latency period in C57BL/6 mice, although this phenomenon was not seen on an outbred model^{12,16,17}.

We initially produced six groups of BRAF^{V600E} inducible ('*Braf*^{CA}') mice representing three pigmentation variants ('black', 'red' and 'albino') with or without epidermal melanocytes (+/- *K14-SCF*, Fig. 1b). Melanocyte-selective expression of BRAF^{V600E} was achieved by tamoxifen-mediated activation of Cre recombinase (*Tyr-Cre(ER)*^{T2}) in adult mice carrying the *Braf*^{CA} allele. The animals were then followed without environmental genotoxic stressors (such as ultraviolet radiation). Black and albino *Braf*^{CA} mice developed similarly low rates of melanoma after a long latency (regardless of *K14-SCF* status). In contrast, red *Braf*^{CA} mice developed melanomas at an accelerated rate with >50% having tumours after 1 year, regardless of *K14-SCF* status (Fig. 1c, d).

The tumours on the black, red and albino backgrounds were grossly amelanotic and largely on the dorsal trunk (within the tamoxifen-treated areas). Occasionally, a tumour would develop on the ventral trunk, tail or paw, which may reflect a predictable spread of tamoxifen secondary to grooming. (Fig. 2a–c). The tumours, which were primarily dermal, were generally amelanotic on the red and albino backgrounds, whereas the melanomas in black mice often had superficial pigmentation adjacent to the epidermis (Fig. 2d–f). Regardless of pigmentation background, the tumours were histologically similar with spindle cell features which were not easily distinguishable from tumours on C57BL/6 *Braf*^{CA}-*Pten*^{flx/flx} animals generated in parallel (compare Fig. 2 with Supplementary Fig. 1). On closer examination, occasional red-BRAF^{V600E} tumour cells were found to contain melanin (Fig. 2g, h). It was further possible to increase pigmentation in the most superficial melanoma cells with topical application of forskolin, an adenylate cyclase agonist known to stimulate skin pigmentation¹⁸ (Fig. 2i). The limited induction of pigmentation is likely related to the poor tissue penetration of forskolin, but nonetheless demonstrates the ability of the melanoma cells to become hyper-pigmented *in vivo* upon activation of cAMP signalling.

Tumours on all three pigmentation backgrounds stained positively for S100, a standard immunohistochemical melanoma marker (Fig. 2j). In addition, reverse transcriptase polymerase chain reaction (RT-PCR) showed that the tumours consistently express the melanocytic pigment genes *M-Mitf*, *Dct*, *Tyrp1* and *Tyr* (Fig. 3a, d and data not shown). In addition, occasional HMB45⁺ cells could be found by immunofluorescence (Supplementary Fig. 2). The tumours on all three pigmentation backgrounds were locally invasive to fat and skeletal muscle with active mitoses. Although no gross visceral organ metastases were observed, small clusters of cells expressing gp100, the pre-melanosome-associated glycoprotein (Pmel/gp100/HMB45), could be found in skin draining lymph nodes (Fig. 2k).

Using a primary cell line derived from one of the red mouse melanomas, we observed that forskolin upregulated the expression of the melanocyte-specific isoform of *Mitf* (*M-Mitf*), and produced a marked increase in expression of the *Dct* and *Tyrp1* pigment genes, consistent with the ability of the cells to respond to melanocytic differentiation signals (Fig. 3a).

To determine whether the melanoma cells were dependent on the presumed oncogenic driver BRAF^{V600E}, we tested their response to small molecule inhibitors of BRAF or MEK (also known as MAP2K). Treatment with the oncogenic BRAF inhibitor, PLX4720, or the MEK inhibitor, U0126, prevented melanoma cell proliferation *in vitro*, and PLX4720 blocked tumour cell growth *in vivo*, consistent with a dependency of these tumours on the BRAF^{V600E} oncoprotein (Fig. 3b, c). BRAF inhibition also elevated the expression of melanocytic genes as previously reported in human melanomas (Fig. 3d)¹⁹.

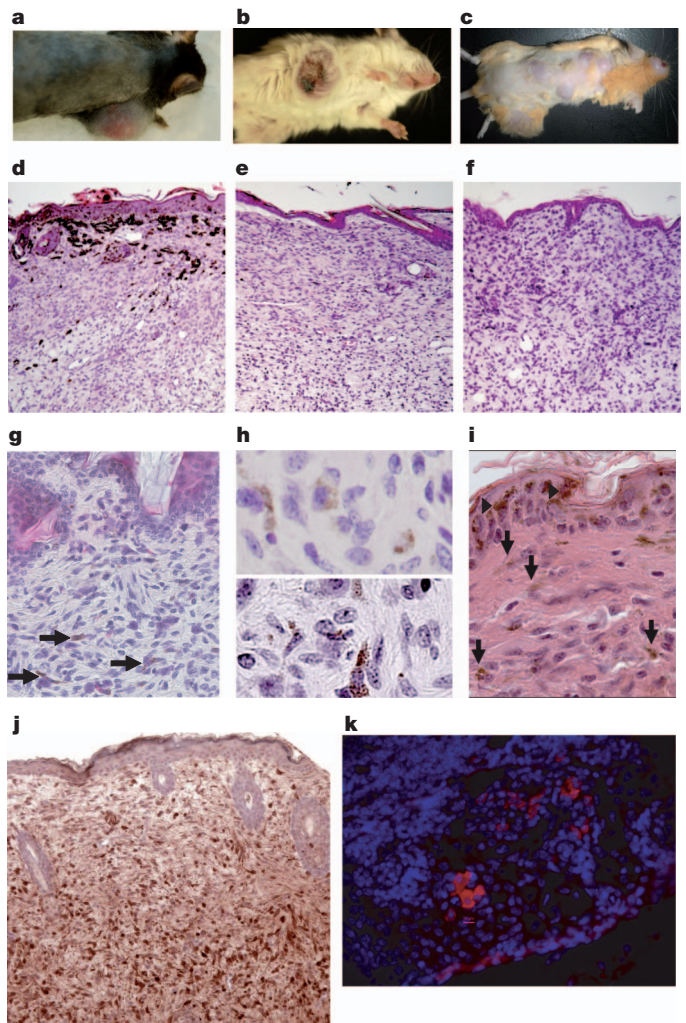


Figure 2 | Melanomas on all three pigmentation variants are morphologically similar and exhibit common histologic features. a–c, Melanomas on black (a), albino (b) and red (c) mice are grossly amelanotic. d–f, Histologically, black (d), albino (e) and red (f) melanomas are also mostly amelanotic, although superficial tumour cells in black-BRAF^{V600E} tumours carry melanin (original magnification, ×10). g, Red-BRAF^{V600E} melanomas can also carry pigment (arrows) (original magnification, ×20). h, Further magnification of two red melanomas also illustrates pigmented tumour cells (original magnification, ×100). i, Forskolin induces epidermal pigmentation (arrowheads) and mild tumour cell pigmentation (arrows) (original magnification, ×20). j, Tumour cells stain positive for S100 (original magnification, ×10). k, Skin-draining lymph nodes carry gp100⁺ cells (red) (scale bar, 10 μm).

Because inactivating mutations in *Mcl1* alter cAMP levels in the cell, red mice undoubtedly have numerous intracellular pathway differences relative to wild-type *Mcl1*^{E/E} (black) animals, including altered DNA repair²⁰. We therefore wished to study whether the pheomelanin pigment pathway itself has an intrinsic mechanistic role, or whether it is merely a marker of melanoma risk. To investigate this question we introduced the albino tyrosinase (*Tyr*^{c/c}) allele into the red *Mcl1*^{E/E} background to test melanoma incidence in albino-*Mcl1*^{E/E} animals, which retain low MC1R activity and also lack all pigment production (Fig. 4a). A melanocyte-targeted *LacZ* transgene was used to confirm that the albino allele does not alter melanocyte number in these mice (Supplementary Fig. 3a, b)¹⁰. As shown in Fig. 4b, the albino allele profoundly protected red mice from melanoma. The rare albino-*Mcl1*^{E/E} melanomas occurred after long latency and had the same amelanotic, S100⁺, histologic features as the other pigmentation variant *Braf*^{CA} animals (Supplementary Fig. 4a–c). This observation suggests

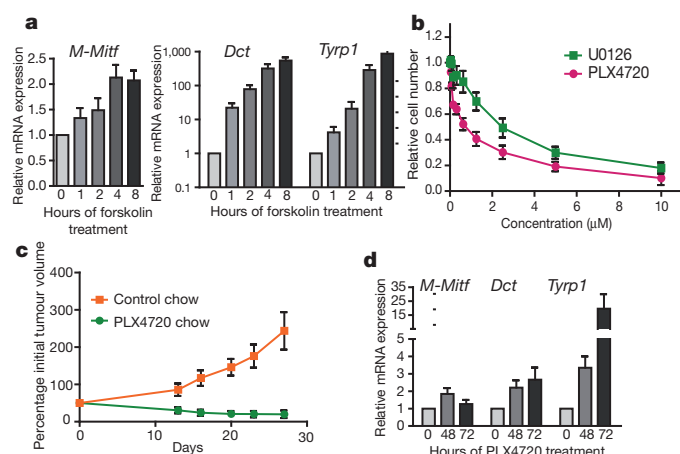


Figure 3 | Tumour cells from a red-*Braf*^{CA} animal behave like classic *BRAF*^{V600E} melanomas after cAMP upregulation or BRAF inhibition. **a**, Forskolin (20 μ M) upregulates expression of melanocytic markers ($n = 4$). **b**, MAPK inhibition by PLX4720 or U0126 decreases melanoma cell proliferation ($n = 3$). **c**, PLX4720 blocks melanoma growth *in vivo* ($n = 3$). **d**, PLX4720 (2 μ M) upregulates expression of melanocytic markers. Relative mRNA expression normalized to 18S ribosomal RNA and 0 h time point. Error bars denote s.e.m.

that the pheomelanin synthesis pathway is necessary for the high rate of ultraviolet-radiation-independent melanoma in the red mice.

Previous studies have demonstrated that ultraviolet radiation amplifies reactive oxygen species (ROS) production and subsequent

oxidative DNA damage in the skin of pigmented mice²¹. Cells with high pheomelanin levels that receive ultraviolet radiation have been found to carry particularly high levels of oxidative damage^{4,5}. Because darkly pigmented individuals carry both pheomelanin and eumelanin, it has been proposed that their lower melanoma risk may result from eumelanin intermediates and polymers absorbing ROS and functioning as *in vivo* antioxidants^{22,23}.

To determine whether ROS-mediated oxidative DNA damage is affected by the pheomelanin synthesis pathway, levels of 8,5'-cyclo-2'-deoxyadenosine (cdA) and 8,5'-cyclo-2'-deoxyguanosine (cdG) were measured in DNA isolated from skin of red-*Mc1r*^{e/e} and albino-*Mc1r*^{e/e} mice, using a previously reported liquid chromatography-tandem mass spectrometric method²⁴ (Fig. 4c). These two ROS-mediated cyclopurines are unlikely to be artificially induced during sample preparation and are quite stable^{25,26}. Significantly, replication studies in *Escherichia coli* have shown that S-cdA and S-cdG can lead to A-to-T and G-to-A mutations at frequencies of 11% and 20%, respectively²⁷. Comparing cyclopurine levels in the skin of various pigmentation-variant mice, it was found that the levels of cdA and cdG are significantly higher in skin from red-*Mc1r*^{e/e} mice compared to skin from albino-*Mc1r*^{e/e} animals (Fig. 4d, e). This observation indicates that activation of the pheomelanin synthesis pathway results in increased oxidative DNA damage. Correlative evidence for increased cellular oxidative stress was also found in the observation that red-*Mc1r*^{e/e} mouse skin carries higher levels of lipid peroxides, a product of ROS-mediated lipid damage (Fig. 4f).

The findings reported here indicate that in the context of oncogenic BRAF activation, individuals carrying red hair/fair skin *MC1R* polymorphisms have an increased risk of melanoma, owing to both poor

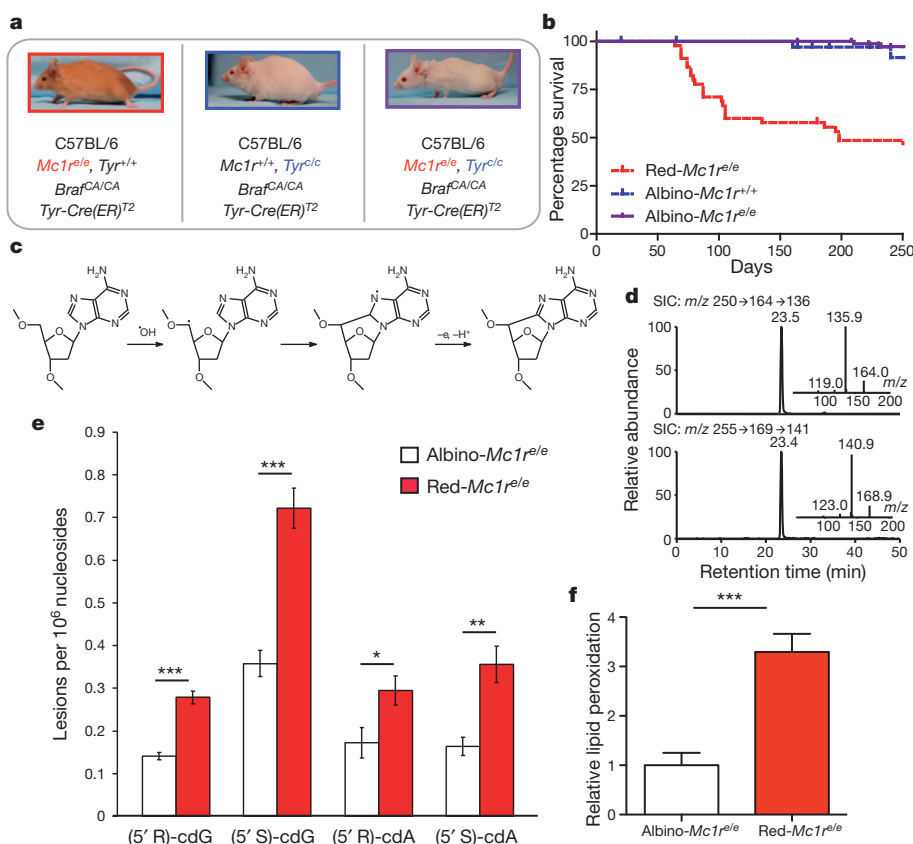


Figure 4 | The ultraviolet-radiation-independent propensity of red *Braf*^{CA} mice to develop melanoma is dependent on pigment production. **a**, Genotypes of mice studied. **b**, The albino allele protects *Mc1r*^{e/e} mice from melanoma development ($n_{\text{red}} = 40$, $n_{\text{albino}} = 48$, $n_{\text{albino-red}} = 90$). $P_{\text{albino-albino-red}} = 0.308$, $P_{\text{albino-red}} < 0.0001$, $P_{\text{red-albino-red}} < 0.0001$. **c**, ROS react with purine nucleosides to produce 8,5'-cyclopurine lesions (cdA shown).

d, Selected-ion chromatograms for DNA from albino-*Mc1r*^{e/e} mouse skin. The insets show the positive-ion MS/MS spectra for unlabelled and labelled S-cdA. **e**, Both diastereomers of cdA and cdG are significantly higher in red-*Mc1r*^{e/e} skin ($n = 3$). * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$. **f**, Lipid peroxide levels are significantly higher in red-*Mc1r*^{e/e} skin ($n = 3$), *** $P < 0.0001$.

protection from environmental carcinogens like ultraviolet radiation, and also via intrinsic carcinogenic features of pheomelanin synthesis; potentially via pheomelanin itself, an intermediate of pigment synthesis or a by-product of the pathway.

In humans, there are multiple *MC1R* polymorphisms with varied perturbation of receptor function that produce a red hair/fair skin phenotype; however, a unifying feature of these various polymorphisms is a high pheomelanin-to-eumelanin ratio, which is also produced by the *Mc1r*^{e/e} allele in mice. Recent data intriguingly demonstrated that black animals overexpressing hepatocyte growth factor (HGF) who receive ultraviolet radiation are at a higher risk of melanoma than their albino counterparts²¹. Although the present study reveals a small difference between black and albino BRAF^{V600E}-driven melanomas in the presence of K14-SCF (Fig. 1d), the effect did not reach statistical significance ($P = 0.103$), perhaps signifying oncogene-specific differences. It seems that the effects of pigmentation and ultraviolet radiation are likely to work together in determining melanoma risk.

The photometer used for our laboratory's routine calibration (International Light 1400) was unable to detect any measurable ultraviolet radiation in our mouse cages during ambient light exposure. However, strong epidemiological work links ultraviolet radiation to melanoma, and the current data do not diminish the importance of sun exposure as a key contributing factor¹. In humans, it is likely that the ultraviolet-independent effects act in concert with ultraviolet-mediated cellular toxicity. In agreement with published studies, ultraviolet radiation at a UV-A/UV-B ratio similar to that found in sunlight (10 J cm⁻² UV-A and 0.65 J cm⁻² UV-B) was found to exacerbate oxidative damage selectively in red mouse skin as measured by levels of lipid peroxidation^{4,5} (Supplementary Fig. 5a). Studies are underway to investigate whether ultraviolet radiation is able to alter the red-BRAF^{V600E} tumour phenotype. Preliminary studies examining the effect of visible light (180 J cm⁻²) did not reveal significantly altered lipid peroxidation in any pigmentation context ($P = 0.4506$). Perhaps, however, there is a trend towards an increased level of lipid peroxidation in red mouse skin (Supplementary Fig. 5b).

Further evidence suggesting an ultraviolet-radiation-independent red hair/fair skin melanoma risk is the observation that although darker-skinned individuals have a significantly lower risk of melanoma than lighter-skinned individuals, the sun protective factor (SPF, a measurement of sunburn protection) of darker skin has been estimated at only in the range of SPF 2.0–4.0 (ref. 28). In addition, sunscreen (typically SPF 20–40) has shown weak efficacy in protecting against melanoma, unlike its protection against cutaneous squamous cell carcinoma^{29,30}. There are numerous potential explanations for the sunscreen-melanoma data including the possibility that ultraviolet radiation shielding may protect against only one of several carcinogenic mechanisms—with the intrinsic pheomelanin pathway representing an additional contributor to melanomagenesis via ultraviolet-radiation-independent means. These data are not evidence against a role for ultraviolet radiation in melanomagenesis. Indeed, the effect of ultraviolet radiation is likely to exacerbate this mechanism, such that ultraviolet radiation shielding and sunscreen remain extremely important for skin cancer prevention. However, further preventative strategies may be essential to optimally diminish melanoma risk in the most susceptible individuals.

METHODS SUMMARY

Mice. At 6–10 weeks of age, mice were treated topically with 20 mg ml⁻¹ tamoxifen for 5 days. For topical darkening, 20% coleus extract was applied as previously described¹⁷. For *in vivo* PLX4720 studies, animals were given *ad libitum* mouse chow containing 2% PLX4720 by weight. All studies and procedures involving animal subjects were approved by the Institutional Animal Care and Use Committees of Massachusetts General Hospital and Dana-Farber Harvard Cancer Center, and were conducted strictly in accordance with the approved animal handling protocol.

Morphological examination. Histology and immunostaining were performed according to established protocols using anti-S100 (Dako), anti-DCT (Santa

Cruz), anti-C5 MITF (tissue culture supernatant), anti-HMB45 (Santa Cruz) and anti-gp100 (Abcam) antibodies.

Primary cell culture. Primary tumour was digested overnight and grown in DMEM. Proliferation was measured by the CellTiter-Glo Luminescent Cell Viability Assay (Promega).

Quantitative RT-PCR. Messenger RNA expression of melanocytic markers was determined using intron-spanning mouse-specific primers with SYBR FAST qPCR master mix (Kapa Biosystems).

Measurement of 8,5'-cyclopurine-2'-deoxynucleosides. cdA and cdG levels were measured as previously published²⁴. Briefly, nuclear DNA was isolated from mouse skin, digested to nucleosides, separated by high-performance liquid chromatography (HPLC) and analysed by liquid chromatography–multi-stage mass spectrometry (LC–MS/MS/MS).

Measurement of lipid peroxidation. Mouse skin was irradiated with ultraviolet (10 J cm⁻² UV-A and 0.65 J cm⁻² UV-B), or visible light (180 J cm⁻²). After homogenization, lipid peroxidation was measured with the OxiSelect TBARS Assay Kit (Cell Biolabs).

Full Methods and any associated references are available in the online version of the paper.

Received 27 February; accepted 27 September 2012.

Published online 31 October 2012.

- Rhodes, A. R., Weinstock, M. A., Fitzpatrick, T. B., Mihm, M. C. J. & Sober, A. J. Risk factors for cutaneous melanoma. A practical method of recognizing predisposed individuals. *J. Am. Med. Assoc.* **258**, 3146–3154 (1987).
- Valverde, P., Healy, E., Jackson, I., Rees, J. L. & Thody, A. J. Variants of the melanocyte-stimulating hormone receptor gene are associated with red hair and fair skin in humans. *Nature Genet.* **11**, 328–330 (1995).
- Rouzaud, F., Kadekaro, A. L., Abdel-Malek, Z. A. & Hearing, V. J. MC1R and the response of melanocytes to ultraviolet radiation. *Mutat. Res.* **571**, 133–152 (2005).
- Wenczl, E. *et al.* (Pheo)melanin photosensitizes UVA-induced DNA damage in cultured human melanocytes. *J. Invest. Dermatol.* **111**, 678–682 (1998).
- Hill, H. Z. & Hill, G. J. UVA, pheomelanin and the carcinogenesis of melanoma. *Pigment Cell Res.* **13** (suppl. 8), 140–144 (2000).
- Curtin, J. A. *et al.* Distinct sets of genetic alterations in melanoma. *N. Engl. J. Med.* **353**, 2135–2147 (2005).
- Elwood, J. M. & Jopson, J. Melanoma and sun exposure: an overview of published studies. *Int. J. Cancer* **73**, 198–203 (1997).
- Robbins, L. S. *et al.* Pigmentation phenotypes of variant extension locus alleles result from point mutations that alter MSH receptor function. *Cell* **72**, 827–834 (1993).
- Halaban, R. *et al.* Tyrosinases of murine melanocytes with mutations at the albino locus. *Proc. Natl Acad. Sci. USA* **85**, 7241–7245 (1988).
- Vanover, J. C. *et al.* Stem cell factor rescues tyrosinase expression and pigmentation in discrete anatomic locations in albino mice. *Pigment Cell Melanoma Res.* **22**, 827–838 (2009).
- Kunisada, T. *et al.* Murine cutaneous mastocytosis and epidermal melanocytosis induced by keratinocyte expression of transgenic stem cell factor. *J. Exp. Med.* **187**, 1565–1573 (1998).
- Dankort, D. *et al.* BRAF^{V600E} cooperates with *Pten* loss to induce metastatic melanoma. *Nature Genet.* **41**, 544–552 (2009).
- Patton, E. E. *et al.* BRAF mutations are sufficient to promote nevi formation and cooperate with p53 in the genesis of melanoma. *Curr. Biol.* **15**, 249–254 (2005).
- Goel, V. K. *et al.* Melanocytic nevus-like hyperplasia and melanoma in transgenic BRAFV600E mice. *Oncogene* **28**, 2289–2298 (2009).
- Michaloglou, C. *et al.* BRAF^{E600E}-associated senescence-like cell cycle arrest of human naevi. *Nature* **436**, 720–724 (2005).
- Dhomen, N. *et al.* Oncogenic Braf induces melanocyte senescence and melanoma in mice. *Cancer Cell* **15**, 294–303 (2009).
- Rae, J. *et al.* BRAF^{V600E}: Tyr-CreERT2::K14-KiIt mice do not develop superficial spreading-like melanoma: keratinocyte kit ligand is insufficient to “translocate” BRAF-driven melanoma to the epidermis. *J. Invest. Dermatol.* **132**, 488–491 (2012).
- D'Orazio, J. A. *et al.* Topical drug rescue strategy and skin protection based on the role of *Mc1r* in UV-induced tanning. *Nature* **443**, 340–344 (2006).
- Boni, A. *et al.* Selective BRAF^{V600E} inhibition enhances T-cell recognition of melanoma without affecting lymphocyte function. *Cancer Res.* **70**, 5213–5219 (2010).
- Kadekaro, A. L. *et al.* Melanocortin 1 receptor genotype: an important determinant of the damage response of melanocytes to ultraviolet radiation. *FASEB J.* **24**, 3850–3860 (2010).
- Noonan, F. P. *et al.* Melanoma induction by ultraviolet A but not ultraviolet B radiation requires melanin pigment. *Nature Commun.* **3**, 884 (2012).
- Nofsinger, J. B., Liu, Y. & Simon, J. D. Aggregation of eumelanin mitigates photogeneration of reactive oxygen species. *Free Radic. Biol. Med.* **52**, 720–730 (2002).
- Kovacs, D. *et al.* The eumelanin intermediate 5,6-dihydroxyindole-2-carboxylic acid is a messenger in the cross-talk among epidermal cells. *J. Invest. Dermatol.* **132**, 1196–1205 (2012).
- Wang, J. *et al.* Quantification of oxidative DNA lesions in tissues of Long-Evans Cinnamon rats by capillary high-performance liquid chromatography–tandem mass spectrometry coupled with stable isotope-dilution method. *Anal. Chem.* **83**, 2201–2209 (2011).

25. Wang, Y. Bulky DNA lesions induced by reactive oxygen species. *Chem. Res. Toxicol.* **21**, 276–281 (2008).
26. Jaruga, P. & Dizdaroglu, M. 8,5'-Cyclopurine-2'-deoxynucleosides in DNA: mechanisms of formation, measurement, repair and biological effects. *DNA Repair (Amst.)* **7**, 1413–1425 (2008).
27. Yuan, B., Wang, J., Cao, H., Sun, R. & Wang, Y. High-throughput analysis of the mutagenic and cytotoxic properties of DNA lesions by next-generation sequencing. *Nucleic Acids Res.* **39**, 5945–5954 (2011).
28. Neugut, A. I., Kizelnik-Freilich, S. & Ackerman, C. Black-white differences in risk for cutaneous, ocular, and visceral melanomas. *Am. J. Public Health* **84**, 1828–1829 (1994).
29. Green, A. C., Williams, G. M., Logan, V. & Strutton, G. M. Reduced melanoma after regular sunscreen use: randomized trial follow-up. *J. Clin. Oncol.* **29**, 257–263 (2011).
30. Huncharek, M. & Kupelnick, B. Use of topical sunscreens and the risk of malignant melanoma: a meta-analysis of 9067 patients from 11 case-control studies. *Am. J. Public Health* **92**, 1173–1177 (2002).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank T. Kunisada for generously sharing *K14-SCF* mice and C. L. Evans for help with mouse skin irradiation. We also thank A. P. Codgill for help with

primary tumour cell culture and A. Piris for pathology consultation as well as M. Haigis and Z. Abdel-Malik for discussions. This work was supported by the following grants from the National Institutes of Health 5R01 AR043369-16 (D.E.F.), R01-CA101864 (Y.W.) and F30 ES020663-01 (D.M.), as well as support from the Dr Miriam and Sheldon G. Adelson Medical Research Foundation, the US-Israel Binational Science Foundation, and the Melanoma Research Alliance (D.E.F.).

Author Contributions D.M. and D.E.F. conceived and planned the project. M.M., M.W.B. and K.M.H. provided the mice carrying the *PTEN*, *BRAF^{V600E}* and *Tyr-Cre(ER)^{Y2}* alleles. D.M. performed the mouse work with help from A.M., J.L., S.P.D. and K.C.R. Histology was performed by D.M., X.L., J.C.V. and J.A.D. with support from D.A.H. Pathological analysis was provided by M.P.H., J.K.L. and M.C.M. *In vitro* studies were performed by D.M. with help from A.M. and J.L. J.A.W. generated the primary mouse cell line. J.W., C.R.G. and Y.W. collected DNA from mouse skin and performed LC-MS/MS/MS. The manuscript was written by D.M. and D.E.F. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.E.F. (dfisher3@partners.org).

METHODS

Mice. All animals used for breeding were backcrossed a minimum of six generations onto the C57BL/6 genetic background (this corresponds to a >98.4% C57BL/6 congenic animal, <http://jaxmice.jax.org/support/nomenclature/tutorial.html>). The black (wild type), red (*Mc1r*^{+/e}) and albino (*Tyr*^{+/c}) animals were purchased from Jackson Laboratories. *K14-SCF* animals were acquired from T. Kunisada. Genotyping of each litter, including the *Tyr-Cre(ER)*^{T2}, *Braf*^{CA} and *PTEN*^{flox/flox} alleles was performed as previously published^{10,11}. At 6–10 weeks of age the dorsal fur was trimmed using animal shears with a 0.25 mm head and the mice were treated topically with 20 mg ml⁻¹ tamoxifen for 5 consecutive days. For tumour darkening, a 20% solution of *Coleus forskohlii* root extract (80 µM forskolin) was topically applied daily as previously described¹⁷. For *in vivo* PLX4720 studies, animals were given *ad libitum* mouse chow containing 2% PLX4720 by weight or control chow acquired from Plexikon. All studies and procedures involving animal subjects were approved by the Institutional Animal Care and Use Committees of Massachusetts General Hospital and Dana-Farber Harvard Cancer Center and were conducted strictly in accordance with the approved animal handling protocol.

Dissection and histology. Tissues of interest were photographed, excised, weighed, rinsed in PBS, fixed in 10% neutral-buffered formalin, rinsed in PBS, and stored in 70% ethanol. Formalin-fixed tissues were paraffin embedded (FFPE) and sectioned (3–5 µm) using standard procedures. Morphological analysis was performed using multiple independent samples per site/organ (5 to 9 samples per genotype) as well as >6 animals. Two pathologists (J.K.L., M.P.H.) independently examined the histopathology of the tumour samples. Digitization and image capture was performed using an Olympus DP70 digital camera (Olympus) connected to an Olympus BX51 light microscope or a Scanscope whole-slide scanning system (Aperio).

Immunohistochemistry. For immunohistochemistry, sections were deparaffinized with xylene and hydrated with a graded series of alcohol. Sections were boiled in 50 mM Tris-buffer (pH 9) or citrate for antigen retrieval and rinsed in PBS. Sections were blocked in 1% BSA, 0.1% Triton X-100 PBS, incubated with 1:200 dilutions of rabbit anti-S100 (Dako), 1:100 dilutions of goat anti-DCT (Santa Cruz), 1:200 dilutions of mouse anti-HMB45 (Santa Cruz) and 1:200 dilutions of mouse anti-gp100 (Abcam) antibodies, followed by visualization with appropriate secondary antibodies conjugated to Alexa594 or Alexa488 (1:500). Appropriate controls for specificity of staining were included and images were captured using an upright fluorescence microscope (Eclipse 90i, Nikon). To identify epidermal melanocytes, skin from reporter mice carrying the various pigmentation alleles and the *K14-SCF* transgene as well as a *DCT-LacZ* reporter allele was cryosectioned and stained with 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal) and nuclear fast red counterstaining.

Primary cell culture. Tumour cells were digested overnight in 10 mg ml⁻¹ collagenase and 1 mg ml⁻¹ hyaluronidase. Initially tumour cells were grown in RPMI media with HEPES and 20% serum. Subsequently tumour cells were grown in DMEM media with 10% serum. Proliferation after 72 h of PLX4720 (Chemietek) and U0126 (Cell Signaling) was determined by the CellTiter-Glo Luminescent Cell Viability Assay (Promega).

Quantitative RT-PCR. RNA was collected from primary cultured tumour cells treated for varying times with forskolin or PLX4720 using the RNeasy Plus mini kit (Qiagen). mRNA expression of melanocytic markers was determined using intron-spanning mouse-specific primers with Kapa SYBR FAST qPCR master mix (Kapa Biosystems). Expression was normalized to 18S rRNA and 0 h time points. Primer sequences used: *Mitf* forward GCCTGAAACCTTGCTATGCTGGAA, *Mitf* reverse AAGGTACTGCTTTACCTGGTGCCT, *Dct* forward AGGTACCATCTGTTGTGGCTGGAA, *Dct* reverse AGTTCCGACTAATCA GCGTTGGGT, *Typr1* forward TGGGGATGTGGATTTCTCTC, *Typr1* reverse AGGGAGAAAGAAGGCTCCTG, *18S* forward AGGTTCTGGCCAACGG TCTAG, *18S* reverse CCCTCTATGGGCAATTTT.

Extraction of nuclear DNA from mouse skin tissues. Nuclear DNA was isolated from mouse skin using a high-salt method. Tissues were ground under liquid nitrogen into fine powder using a mortar and pestle. A nuclei lysis buffer containing 20 mM Tris (pH 8.3), 20 mM EDTA, 400 mM NaCl, 1% SDS (w/v) and 0.05% proteinase K (w/v) was added to the tissue and incubated in a water bath at 55 °C overnight. Half volume of saturated NaCl solution was added to the digestion mixture, incubated at 55 °C for 15 min then centrifuged at ~12,000g

for 30 min. The supernatant was collected and centrifuged again. The nucleic acids in the supernatant were precipitated with cold ethanol, dissolved in water and incubated in the presence of 0.03% RNase A (w/v) and 0.25 U µl⁻¹ of RNase T1 at 37 °C overnight, and subsequently extracted with an equal volume of chloroform/isoamyl alcohol (24:1, v/v) twice. The DNA was then precipitated from the aqueous layer by cold ethanol, centrifuged at 10,000g at 4 °C for 15 min, washed twice with 70% cold ethanol and dried under vacuum. The DNA pellet was dissolved in deionized water and quantified by using ultraviolet absorption spectrophotometry.

Enzymatic digestion of nuclear DNA. Nuclease P1 (16 U), phosphodiesterase 2 (0.025 U), 20 nmol of erythro-9-(2-hydroxy-3-nonyl) adenine (EHNA) and a 30 µl solution containing 300 mM sodium acetate (pH 5.6) and 10 mM zinc chloride were added to 200 µg of DNA. In this context, EHNA served as an inhibitor for deamination of 2'-deoxyadenosine to 2'-deoxyinosine (dI) induced by adenine deaminase. The above digestion was continued at 37 °C for 48 h. To the digestion mixture were then added alkaline phosphatase (10 U), phosphodiesterase 1 (0.0125 U) and 60 µl of 0.5 M Tris-HCl buffer (pH 8.9). The digestion was continued at 37 °C for 2 h and subsequently neutralized by addition of formic acid. To the mixture were then added uniformly ¹⁵N-labelled standard lesions, which included 400 fmol of R-cdG, 150 fmol of S-cdG, 80 fmol of R-cdA and 40 fmol of S-cdA. The enzymes in the digestion mixture were subsequently removed by chloroform extraction twice. The resulting aqueous layer was subjected to off-line high-performance liquid chromatography (HPLC) separation for the enrichment of the lesions under study, following our previously described procedures²⁶.

Liquid chromatography–multi-stage mass spectrometry (LC–MS/MS/MS) analysis. The LC–MS/MS/MS experiments were conducted using an LTQ linear ion trap mass spectrometer using our recently described conditions²⁶. Briefly, the amounts of cdA and cdG lesions in each nucleoside sample were calculated based on the ratios of peak areas found in the selected-ion chromatograms for the analyte (for example, the 23.5 min peak in the top panel of Fig. 4d for S-cdA) and the corresponding stable isotope-labelled standard (for example, the 23.4 min peak in the bottom panel of Fig. 4d for the ¹⁵N-labelled S-cdA), the known amount of uniformly ¹⁵N-labelled standard was added to the nucleoside mixture (for example, 40 fmol for S-cdA), and calibration curves. The calibration curves were constructed from the same LC–MS/MS/MS analyses of a series of mixtures with known compositions of the unlabelled cdA, cdG and constant amounts of the corresponding uniformly ¹⁵N-labelled standards, as described previously²⁶. The lesion formation frequencies as shown in Fig. 4e were then calculated by dividing the amounts of cdA and cdG in the sample with the total amount of nucleosides present.

Skin irradiation and lipid peroxide measurement. Six-week-old mice were euthanized and fur was removed using animal shears with a 0.25 mm head. Twelve sections of skin, each with an area of 1 cm², were removed from each mouse and placed in 35-mm dishes on ice after adherence to Whatman filter paper suspended in PBS. For each ultraviolet and visible light study, six sections of skin from each mouse were placed in the dark on ice as controls. For ultraviolet studies, six sections of skin from each mouse were irradiated on ice with 10 J cm⁻² UV-A and 0.65 J cm⁻² UV-B at an irradiance of 6.67 mW cm⁻² using a Sylvania 350 Blacklight (Osram Sylvania). This ultraviolet distribution is comparable to natural sunlight (96.65% UV-A and 3.35% UV-B). Two mice of each pigmentation type were used for a total of *n* = 12 skin samples for each condition. For visible light studies, six sections of skin from each mouse were irradiated on ice with 180 J cm⁻² visible light from a Dolan-Jenner A3200 Fiber-Lite Illuminator at an irradiance of 200 mW cm⁻². The illuminator bulb was fit with a Thorlabs FEL0400 Edgepass ultraviolet filter with a transmission of <0.001% for wavelengths <400 nm, such that no irradiation output was detectable in the ultraviolet range below 400 nm. One mouse of each pigmentation type was used for a total of six skin samples for each condition. Following treatment, skin sections were flash-frozen and homogenized in PBS containing the antioxidant butylated hydroxytoluene (BHT) to prevent further lipid peroxidation, using a Qiagen TissueLyser II. Homogenized samples were centrifuged and supernatants were collected. Protein content of each sample was determined by Coomassie Plus Protein Assay, and samples were diluted with PBS plus BHT for normalization of sample concentration (Thermo Scientific). Lipid peroxidation of each irradiated set of sample was determined using an OxiSelect TBARS Assay kit and normalized to its unirradiated control (Cell Biolabs).

Long non-coding antisense RNA controls *Uchl1* translation through an embedded SINEB2 repeat

Claudia Carrieri^{1*}, Laura Cimatti^{1*}, Marta Biagioli^{1,2}, Anne Beugnet³, Silvia Zucchelli^{1,2}, Stefania Fedele¹, Elisa Pesce³, Isidre Ferrer⁴, Licio Collavin^{5,6}, Claudio Santoro⁷, Alistair R. R. Forrest⁸, Piero Carninci⁸, Stefano Biffo^{3,9}, Elia Stupka¹⁰ & Stefano Gustincich^{1,2}

Most of the mammalian genome is transcribed^{1–3}. This generates a vast repertoire of transcripts that includes protein-coding messenger RNAs, long non-coding RNAs (lncRNAs) and repetitive sequences, such as SINEs (short interspersed nuclear elements). A large percentage of ncRNAs are nuclear-enriched with unknown function⁴. Antisense lncRNAs may form sense-antisense pairs by pairing with a protein-coding gene on the opposite strand to regulate epigenetic silencing, transcription and mRNA stability^{5–10}. Here we identify a nuclear-enriched lncRNA antisense to mouse ubiquitin carboxy-terminal hydrolase L1 (*Uchl1*), a gene involved in brain function and neurodegenerative diseases¹¹. Antisense *Uchl1* increases UCHL1 protein synthesis at a post-transcriptional level, hereby identifying a new functional class of lncRNAs. Antisense *Uchl1* activity depends on the presence of a 5' overlapping sequence and an embedded inverted SINEB2 element. These features are shared by other natural antisense transcripts and can confer regulatory activity to an artificial antisense to green fluorescent protein. Antisense *Uchl1* function is under the control of stress signalling pathways, as mTORC1 inhibition by rapamycin causes an increase in UCHL1 protein that is associated to the shuttling of antisense *Uchl1* RNA from the nucleus to the cytoplasm. Antisense *Uchl1* RNA is then required for the association of the overlapping sense protein-coding mRNA to active polysomes for translation. These data reveal another layer of gene expression control at the post-transcriptional level.

To discover non-coding antisense transcripts of sense-antisense (S-AS) pairs expressed in the brain, the mouse syntenic loci of genes involved in neurodegenerative diseases were identified computationally and examined in the Ensembl browser (<http://www.ensembl.org>). The FANTOM2 clone 6430596G22 was classified as a spliced antisense lncRNA of the *Uchl1* gene¹¹; we refer to this as antisense *Uchl1*. UCHL1 is a neuron-restricted protein that acts as a deubiquitinating enzyme, ubiquitin ligase or monoubiquitin stabilizer¹². An in-frame deletion in the *Uchl1* gene, as in gracile axonal dystrophy mice, leads to ataxia and axonal degeneration. Although an association of UCHL1 gene mutations to familial Parkinson's disease has not been confirmed in independent families, oxidative inactivation of UCHL1 protein has been reported in Parkinson's disease and Alzheimer's disease brains^{13–15}.

Antisense *Uchl1* is a 5' head-to-head transcript that initiates within the second intron of *Uchl1* and overlaps the first 73 nucleotides of the sense mRNA including the AUG codon. By 5' rapid amplification of cDNA ends (RACE), the transcriptional start site (TSS) of antisense *Uchl1* was mapped to the second intron of *Uchl1* (Fig. 1a). The non-overlapping part of the transcript contains two embedded repetitive sequences, SINEB1 of the F1 subclass (Alu) and SINEB2 of the B3 subclass, identified by Repeatmasker^{16–18}. The FANTOM2 clone spans

a genomic region of 70 kilobases (kb) (Fig. 1a) and its genomic organization is conserved in mammals (Supplementary Fig. 1a).

Sense and antisense *Uchl1* expression in mouse and human tissues was similar (Fig. 1b and Supplementary Fig. 1b). In the mouse, antisense *Uchl1* RNA was highly expressed in the ventral midbrain (Fig. 1b) and in the MN9D dopaminergic cell line (data not shown). Mature *Uchl1* mRNA was observed mainly in the cytoplasm of dopaminergic neurons, whereas antisense *Uchl1* was enriched in the nucleus (Fig. 1c and Supplementary Fig. 2). Antisense *Uchl1* expression was confirmed by qRT-PCR from dopaminergic neurons purified with laser capture microdissection (LCM, Supplementary Fig. 3).

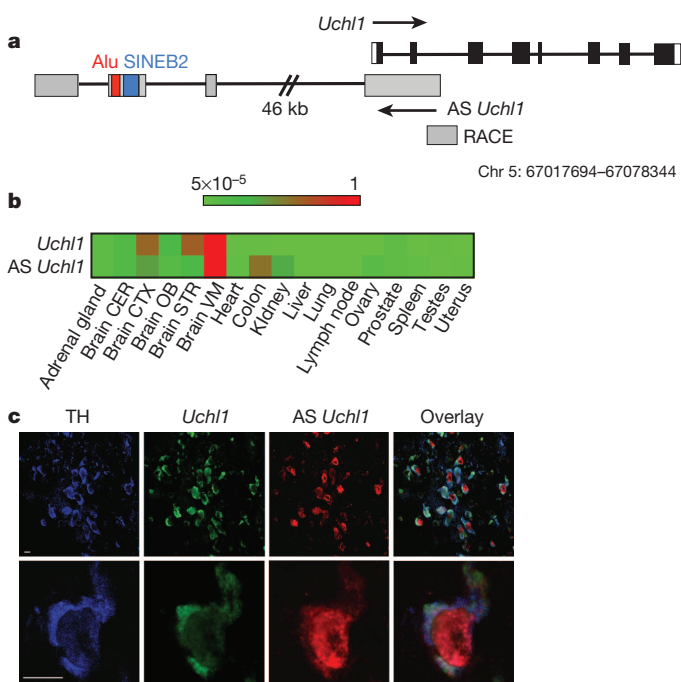


Figure 1 | Expression of antisense *Uchl1* in dopaminergic neurons. **a**, *Uchl1*/antisense (AS) *Uchl1* genomic organization. *Uchl1* exons are in black; 3' and 5' UTRs are in white; antisense *Uchl1* exons are grey; repetitive elements are in red (Alu) and blue (SINEB2). Introns are indicated as lines. **b**, Quantitative expression of *Uchl1* and antisense *Uchl1* in mouse tissues ($\Delta\Delta Ct/\Delta\Delta Ct_{max}$). CER, cerebellum; CTX, cortex; OB, olfactory bulb; STR, striatum; VM, ventral midbrain. **c**, Antisense *Uchl1* (red) and *Uchl1* (green) transcripts are expressed in the nucleus and cytoplasm of TH-positive dopaminergic neurons of the substantia nigra (blue).

¹Area of Neuroscience, International School for Advanced Studies (SISSA), via Bonomea 265, 34136 Trieste, Italy. ²The Giovanni Armenise-Harvard Foundation Laboratory, via Bonomea 265, 34136 Trieste, Italy. ³Laboratory of Molecular Histology and Cell Growth, DIBIT, San Raffaele Scientific Institute, Via Olgettina 58, 20132 Milan, Italy. ⁴Institute of Neuropathology, IDIBELL-University Hospital of Bellvitge, Carrer Feixa Llarga s/n, 08907 Hospitalet de Llobregat, Spain. ⁵Laboratorio Nazionale Consorzio Interuniversitario Biotecnologie (LNCIB), Area Science Park, Padriciano 99, 34149 Trieste, Italy. ⁶Department of Life Sciences (DSV), University of Trieste, via Giorgieri 1, 34129 Trieste, Italy. ⁷Department of Health Sciences, University of Eastern Piedmont, Viale Solaroli 17, 28100 Novara, Italy. ⁸Omics Science Center, RIKEN Yokohama Institute, 1-7-22 Suehiro-chô, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan. ⁹Department of Environmental and Life Sciences, University of Eastern Piedmont, Viale T. Michel 11, 15121 Alessandria, Italy. ¹⁰Center for Translational Genomics and Bioinformatics, San Raffaele Scientific Institute, Via Olgettina 58, 20132 Milan, Italy.

*These authors contributed equally to this work.

Transient expression of antisense *Uchl1* in MN9D cells caused no significant change in endogenous *Uchl1* mRNA expression. Notably, a strong and reproducible upregulation of UCHL1 protein was detected within 24 h (Fig. 2a). When increasing amounts of antisense *Uchl1* were co-transfected with murine *Uchl1* into HEK cells, which do not express either transcript, dose-dependent UCHL1 protein upregulation was recorded in the absence of any significant change in the quantity of exogenous *Uchl1* mRNA (Fig. 2b). These data indicate that antisense *Uchl1* regulates UCHL1 expression at a post-transcriptional level.

Antisense *Uchl1* deletion constructs lacking the 5' first exon (antisense *Uchl1*($\Delta 5'$)) or the last three exons (antisense *Uchl1*($\Delta 3'$)) failed to

induce UCHL1 protein in MN9D and HEK cells, suggesting that both 5' and 3' components are important to antisense *Uchl1* function (Fig. 2c and Supplementary Fig. 4a). Targeted deletion of the region containing the embedded SINEB2 and Alu repetitive sequences (Δ AS) was also able to prevent UCHL1 protein induction. Deletion of each repetitive element separately revealed that SINEB2 is the functional unit required by antisense *Uchl1* for increasing UCHL1 protein synthesis (Fig. 2d). In all cases no change in *Uchl1* mRNA level was detected (Supplementary Fig. 4b). A mutant with a flipped SINEB2 sequence was unable to increase UCHL1 protein levels, thus proving the orientation-dependent activity of the SINEB2 domain embedded within antisense *Uchl1* (Fig. 2d). Importantly, an artificial construct containing the 73-nucleotide overlapping sequence immediately close to the repetitive elements in antisense *Uchl1*($\Delta 5'$) increased UCHL1 levels as much as the full-length clone (Fig. 2e and Supplementary Fig. 4c).

We then considered whether other SINEB2-containing lncRNAs may post-transcriptionally regulate the expression of their protein-coding partner, on the basis of similar structural elements. The FANTOM3 collection of non-coding cDNAs was bioinformatically screened for natural antisense transcripts that contain SINEB2 elements of the B3 subclass in the correct orientation and 5' head-to-head overlapping to a protein-coding gene. This identified 31 S-AS pairs similar to the *Uchl1*/antisense *Uchl1* structure (Supplementary Fig. 5). By sequence alignment, we chose antisense *Uxt* (4833404H03), antisense of ubiquitously expressed transcript (*Uxt*), as the one with the most similar SINEB2 elements (Fig. 3a). Transfection of antisense *Uxt* in MN9D cells elicited an increase of UXT protein level with no change in *Uxt* mRNA (Fig. 3b), indicating that antisense *Uxt* was similarly able to increase protein levels post-transcriptionally.

These data indicate a model whereby lncRNAs regulate protein synthesis through the combined activities of two domains. The 5' antisense

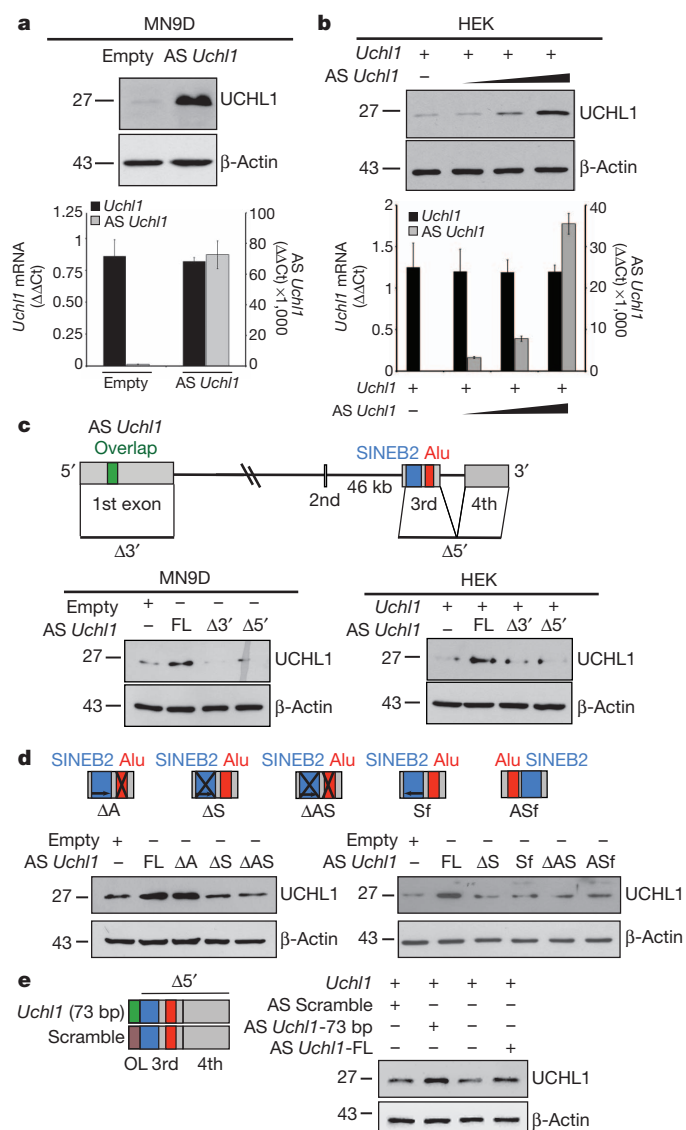


Figure 2 | Antisense *Uchl1* regulates UCHL1 protein levels via an embedded inverted SINEB2 element. **a**, Antisense *Uchl1*-transfected dopaminergic MN9D cells show increased levels of endogenous UCHL1 protein, with unchanged mRNA quantity. **b**, Increasing doses of transfected antisense *Uchl1* titrate UCHL1 protein but not mRNA levels in HEK cells. Data in **a** and **b** indicate mean \pm s.d., $n \geq 3$. **c**, Full-length (FL) antisense *Uchl1* is required for regulating endogenous (MN9D cells, left panel) and overexpressed (HEK cells, right panel) UCHL1 protein levels. Scheme of $\Delta 5'$ or $\Delta 3'$ deletion mutants is shown. **d**, Inverted SINEB2 is sufficient to control endogenous UCHL1 protein levels in MN9D cells. Scheme of mutants is shown in 5' to 3' orientation. ΔA , Δ Alu; ΔS , Δ SINEB2; ΔAS , Δ Alu + SINEB2; Sf, SINEB2 flipped; ASf, Alu + SINEB2 flipped. **e**, A 73-bp overlap (OL) of antisense *Uchl1* is sufficient to increase UCHL1 in transfected HEK cells. Scheme of mutant and scramble control in 5' to 3' orientation. Units for numbers along the left of gels in **a–e** indicate kDa.

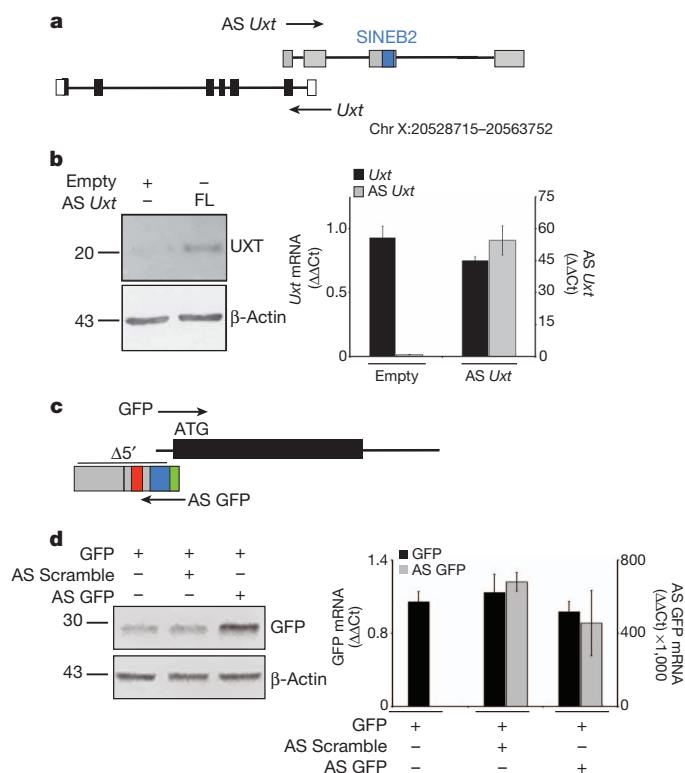


Figure 3 | Natural and synthetic antisense lncRNAs increase target protein levels. **a**, Scheme of *Uxt*/antisense *Uxt* genomic organization. **b**, Antisense *Uxt* increases endogenous UXT protein levels (left) without affecting RNA levels (right) in transfected MN9D cells. **c**, Scheme of antisense GFP construct. $\Delta 5'$ with repetitive elements (SINEB2, blue; Alu, red) and overlap (green) regions is indicated. **d**, Inverted SINEB2 plus the overlap sequence increase GFP levels in transfected cells. Data in **b** and **d** indicate mean \pm s.d., $n \geq 3$.

region provides specificity for the sense target gene whereas the repetitive element confers the protein synthesis activation domain. The model predicts that by swapping the overlapping sequence one may increase the amount of proteins encoded by the mRNAs of choice acting at the post-transcriptional level. We thus synthesized a 72-nucleotide-long artificial sequence antisense to the AUG-containing region as transcribed from pEGFP, and inserted it into antisense *Uchl1*($\Delta 5'$) to generate antisense GFP (Fig. 3c). Co-transfection of antisense GFP with pEGFP strongly increased GFP protein but not mRNA levels in HEK cells (Fig. 3d). When we pulsed cells with methionine for an hour and immunoprecipitated GFP, antisense GFP induced an increase in radioactively labelled, neo-synthesized GFP, without affecting mRNA levels (Supplementary Fig. 6).

To understand how the antisense *Uchl1* transcript operates and the physiological conditions in which it might act, we assayed several stimuli and/or drugs for their ability to modulate UCHL1 protein expression. Inhibition of mTORC1 signalling favoured an increase in UCHL1 levels in a range from 1.5- to 2.5-fold (Fig. 4a). This effect was evident with as low as 20 nM rapamycin (Supplementary Fig. 7) and was concomitant with dephosphorylation of mTOR targets p70S6K and 4E-BP1. Furthermore, the effect was not due to a stabilization of the protein, as co-application of cycloheximide decreased UCHL1 protein levels (Supplementary Fig. 8a). These data are surprising because rapamycin impairs formation of the CAP-dependent complex and hence translation of highly structured mRNAs¹⁹. However, in agreement with previous reports, in our experimental settings mTORC1 inhibition only slightly impairs the global rate of translation (Supplementary Fig. 8b). Under these conditions, it has been proposed that rapamycin may affect competition among different mRNAs²⁰. If so, we proposed that inhibition of the CAP complex formation favours the translation of *Uchl1* mRNA with a mechanism that requires antisense *Uchl1*.

To test this model, we used two complementary approaches to establish a loss-of-function phenotype. First, we downregulated antisense *Uchl1* levels with short hairpin RNA (shRNA) targeting its promoter region. MN9D cells constitutively expressing shRNA for antisense *Uchl1* did not show any changes in UCHL1 protein levels upon rapamycin treatment, whereas scramble control cells showed UCHL1 upregulation as in the parental line (Fig. 4b). Dephosphorylation of p70S6K and 4E-BP1 proved that rapamycin inhibited mTOR activity as expected. We then exploited the dominant-negative property of an antisense *Uchl1* mutant lacking the SINEB2 repeat element. Overexpression of antisense *Uchl1*(Δ SINEB2) inhibited the ability of full-length antisense *Uchl1* to increase protein levels (Supplementary Fig. 9). Upon rapamycin treatment, cells stably expressing antisense *Uchl1*(Δ SINEB2) did not show any UCHL1 protein induction despite dephosphorylation of mTOR targets (Fig. 4c). These complementary experiments prove that functional antisense expression is required for UCHL1 protein increase elicited by rapamycin.

Because antisense *Uchl1* transcript is enriched in the nucleus of dopaminergic neurons, we measured antisense *Uchl1* and *Uchl1* RNA content in the nucleus and cytoplasm of MN9D cells upon rapamycin treatment. As shown in Fig. 4d, rapamycin substantially increased antisense *Uchl1* concentration in the cytoplasmic fraction. This effect was confirmed by a concomitant decrease in its nuclear steady-state levels, and by the absence of any *de novo* transcription. The total content of primary and spliced transcripts remained constant (data not shown). *Uchl1* mRNA showed no change in subcellular distribution, *de novo* transcription or total cellular content. These data demonstrate that antisense *Uchl1* localization can be regulated by the mTOR pathway, and its cytoplasmic level correlates with the expression of UCHL1 protein.

We therefore monitored the association of *Uchl1* mRNA with polysomes to assess the role of translation in antisense *Uchl1*-mediated UCHL1 protein induction. Fractionated MN9D cell extracts were used to measure the recruitment of *Uchl1* mRNA on polysomes by

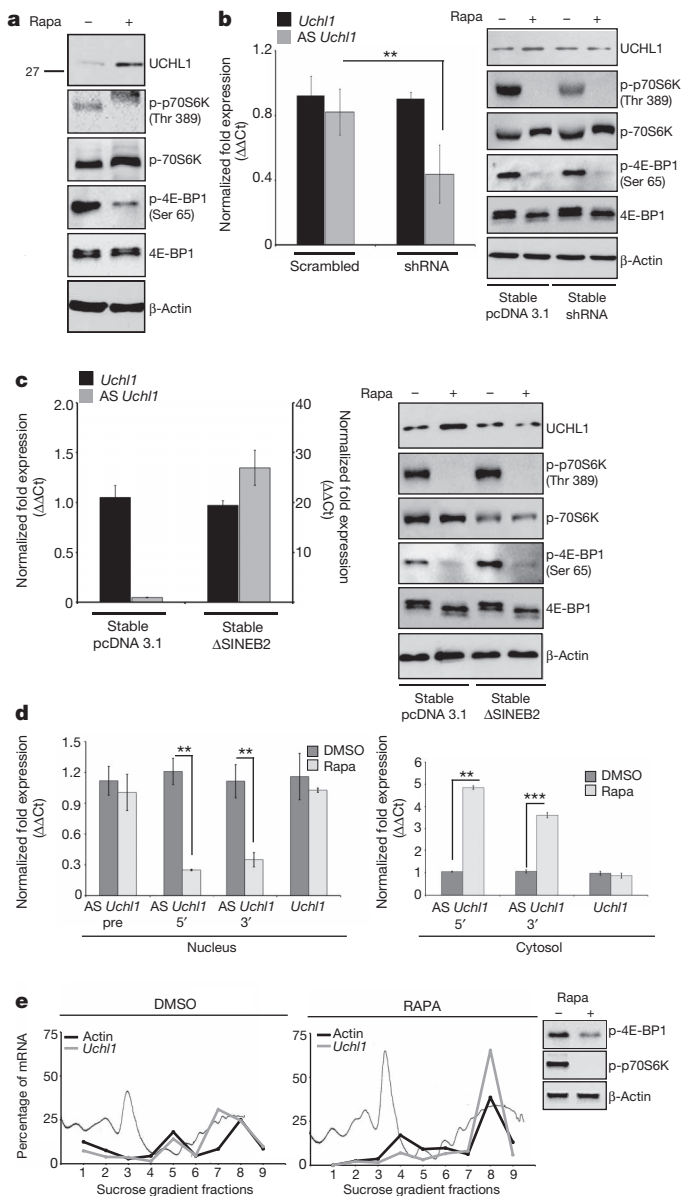


Figure 4 | Antisense *Uchl1* mediates UCHL1 protein induction by rapamycin. **a**, UCHL1 protein level is increased in rapamycin-treated MN9D cells. Rapamycin inhibition of mTOR pathway is verified with anti-p-p70S6K and anti-p-4E-BP1 antibodies. **b**, Silencing antisense *Uchl1* (shRNA) in MN9D cells inhibits rapamycin-induced UCHL1 protein level. Left, mRNA levels; right, protein levels. **c**, Deletion of embedded SINEB2 (Δ SINEB2) is sufficient to inhibit rapamycin-induced UCHL1 protein upregulation. **d**, Antisense *Uchl1* translocates to the cytoplasm upon rapamycin treatment in MN9D cells. mRNA levels were measured with 5' or 3' primers. Data in **b–d** indicate mean \pm s.d., $n \geq 3$. ** $P < 0.01$; *** $P < 0.005$. **e**, Rapamycin increases *Uchl1* mRNA in heavy polysomes; the absorbance profile is outlined in the background of each plot.

qRT-PCR and northern blotting (Fig. 4e and Supplementary Fig. 10). In basal conditions, *Uchl1* mRNA was associated with translating ribosomes. Rapamycin treatment induced a shift of *Uchl1* mRNA to heavier polysomes, consistent with an enhanced rate of translation initiation; this increase of *Uchl1* mRNA association to heavier polysomes did not occur in cells overexpressing the dominant-negative form of antisense *Uchl1* (Supplementary Fig. 11).

Antisense *Uchl1* is the representative member of a new functional class of lncRNAs that are part of S-AS pairs in the mammalian genome that require overlap at the 5' end and the action of a SINEB2 repeat. This new function for SINEB2 sequences in the cytoplasm adds to their

well-established role in the nucleus as inhibitors of RNA polymerase II¹⁶. Stress-dependent nucleocytoplasmic shuttling of lncRNAs may be a common strategy to regulate translation, as CTN-RNA, another nuclear-retained lncRNA, was found to have a cryptic protein-coding sequence at its 3' end when in the cytoplasm²¹.

It is intriguing that this nuclear lncRNA-mediated mechanism for post-transcriptional control of gene expression is active when CAP-dependent translation is inhibited by rapamycin. This drug blocks mTORC1 kinase, which activates the eIF4F complex^{19,22}. However, some mRNAs escape mTORC1 inhibition by being able to be recruited to ribosomes in an eIF4F-independent manner for presenting complex mRNA loops that function as internal ribosomal entry sites (IRES)²³. Indeed, IRES-mediated translation is prominent in conditions of stress or growth factor inhibition and its alteration affects processes such as tumorigenesis^{22,24}. In genetic and neurochemical models of Parkinson's disease, mTORC1 inhibition protects dopaminergic neurons from apoptosis^{25,26}.

Antisense lncRNA-mediated translation may be another mechanism to maintain synthesis of pro-survival proteins, such as UCHL1, that are involved in rapamycin neuroprotective function and more generally in cellular response to stress. This mechanism may represent the outcome of an evolutionary pressure on the genomic organization of anti-stress elements to favour gene-specific regulation of translation when CAP-dependent initiation is reduced. Finally, natural and synthetic antisense transcripts with embedded repetitive elements may represent molecular tools to increase translation of selected mRNAs, defining a potential new class of RNA therapeutics.

METHODS SUMMARY

5' RACE for antisense *Uchl1* was performed with Gene Racer (Invitrogen). Double-fluorescence *in situ* hybridization of biotin- and digoxigenin-labelled probes was detected using fluorochrome-conjugated reagents. Images were captured with Confocal Laser Microscopy (LEICA). Expression of antisense *Uchl1* was performed on neuronal cell lines and dopaminergic neurons purified with LCM from TH-GFP mice. Cell lines were cultured under standard conditions. shRNA targeting the promoter of antisense *Uchl1* was cloned in pSUPERIOR.Neo.GFP vector (Oligo Engine). Polysomes were prepared by sucrose gradient and associated *Uchl1* mRNA was measured by qRT-PCR and northern blotting.

Full Methods and any associated references are available in the online version of the paper.

Received 3 December 2010; accepted 14 August 2012.

Published online 14 October 2012.

1. Birney, E. *et al.* Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799–816 (2007).
2. The FANTOM Consortium. The transcriptional landscape of the mammalian genome. *Science* **309**, 1559–1563 (2005).
3. Kapranov, P., Willingham, A. T. & Gingeras, T. R. Genome-wide transcription and the implications for genomic organization. *Nature Rev. Genet.* **8**, 413–423 (2007).
4. Kapranov, P. *et al.* RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* **316**, 1484–1488 (2007).
5. Beltran, M. *et al.* A natural antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial-mesenchymal transition. *Genes Dev.* **22**, 756–769 (2008).
6. Ebralidze, A. K. *et al.* PU.1 expression is modulated by the balance of functional sense and antisense RNAs regulated by a shared cis-regulatory element. *Genes Dev.* **22**, 2085–2092 (2008).
7. Hastings, M. L., Ingle, H. A., Lazar, M. A. & Munroe, S. H. Post-transcriptional regulation of thyroid hormone receptor expression by cis-acting sequences and a naturally occurring antisense RNA. *J. Biol. Chem.* **275**, 11507–11513 (2000).
8. Huarte, M. *et al.* A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell* **142**, 409–419 (2010).

9. Katayama, S. *et al.* Antisense transcription in the mammalian transcriptome. *Science* **309**, 1564–1566 (2005).
10. Spigoni, G., Gedressi, C. & Mallamaci, A. Regulation of *Emx2* expression by antisense transcripts in murine cortico-cerebral precursors. *PLoS ONE* **5**, e8658 (2010).
11. Setsue, R. & Wada, K. The functions of UCH-L1 and its relation to neurodegenerative diseases. *Neurochem. Int.* **51**, 105–111 (2007).
12. Liu, Y., Fallon, L., Lashuel, H. A., Liu, Z. & Lansbury, P. T. Jr. The UCH-L1 gene encodes two opposing enzymatic activities that affect α -synuclein degradation and Parkinson's disease susceptibility. *Cell* **111**, 209–218 (2002).
13. Barrachina, M. *et al.* Reduced ubiquitin C-terminal hydrolase-1 expression levels in dementia with Lewy bodies. *Neurobiol. Dis.* **22**, 265–273 (2006).
14. Barrachina, M. *et al.* Amyloid- β deposition in the cerebral cortex in dementia with Lewy bodies is accompanied by a relative increase in A β PP mRNA isoforms containing the Kunitz protease inhibitor. *Neurochem. Int.* **46**, 253–260 (2005).
15. Choi, J. *et al.* Oxidative modifications and down-regulation of ubiquitin carboxyl-terminal hydrolase L1 associated with idiopathic Parkinson's and Alzheimer's diseases. *J. Biol. Chem.* **279**, 13256–13264 (2004).
16. Nishihara, H., Smit, A. F. & Okada, N. Functional noncoding sequences derived from SINEs in the mammalian genome. *Genome Res.* **16**, 864–874 (2006).
17. Ponicsan, S. L., Kugel, J. F. & Goodrich, J. A. Genomic gems: SINE RNAs regulate mRNA production. *Curr. Opin. Genet. Dev.* **20**, 149–155 (2010).
18. Quentin, Y. Fusion of a free left Alu monomer and a free right Alu monomer at the origin of the Alu family in the primate genomes. *Nucleic Acids Res.* **20**, 487–493 (1992).
19. Andrei, M. A. *et al.* A role for eIF4E and eIF4E-transporter in targeting mRNPs to mammalian processing bodies. *RNA* **11**, 717–727 (2005).
20. Merrick, W. C. Eukaryotic protein synthesis: still a mystery. *J. Biol. Chem.* **285**, 21197–21201 (2010).
21. Prasanth, K. V. *et al.* Regulating gene expression through RNA nuclear retention. *Cell* **123**, 249–263 (2005).
22. Holcik, M. & Sonenberg, N. Translational control in stress and apoptosis. *Nature Rev. Mol. Cell Biol.* **6**, 318–327 (2005).
23. Gilbert, W. V. Alternative ways to think about cellular internal ribosome entry. *J. Biol. Chem.* **285**, 29033–29038 (2010).
24. Yoon, A. *et al.* Impaired control of IRES-mediated translation in X-linked dyskeratosis congenita. *Science* **312**, 902–906 (2006).
25. Malagelada, C., Jin, Z. H., Jackson-Lewis, V., Przedborski, S. & Greene, L. A. Rapamycin protects against neuron death in *in vitro* and *in vivo* models of Parkinson's disease. *J. Neurosci.* **30**, 1166–1175 (2010).
26. Santini, E., Heiman, M., Greengard, P., Valjent, E. & Fisone, G. Inhibition of mTOR signaling in Parkinson's disease prevents L-DOPA-induced dyskinesia. *Sci. Signal.* **2**, ra36 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank S.G. laboratory members for thought-provoking discussions and C. Leonesi for technical help. We thank F. Persichetti, A. Mallamaci, E. Calautti, S. Saoncella, A. Lunardi, D. De Pietri Tonelli, R. Sanges, M. E. MacDonald and T. Perlmann for support and discussions; and M. J. Zigmond and B. Joseph for sharing the MN9D cell line. This work was supported by the FP7 Dopaminet to S.G., E.S. and P.C., by The Giovanni Armenise-Harvard Foundation to S.G. and by the Compagnia di San Paolo to S.B.

Author Contributions C.C. designed and performed the experiments, and analysed the results; L.Ci. designed and performed the experiments, and analysed the results; M.B. designed and performed the experiments, and analysed the results; A.B. prepared polysomes; S.Z. designed the experiments, analysed the results and wrote the manuscript; S.F. carried out qRT-PCR on polysome fractions and the pulse labelling experiment; E.P. prepared polysomes and carried out northern blotting; I.F. analysed the results; L.Co. designed the experiments and analysed the results; C.S. analysed the data and discussed the results; A.R.R.F. performed bioinformatic analysis for the identification of SINEB2 and family members and designed Δ Alu and Δ SINEB2 mutants; P.C. provided reagents, experimental design and managing; S.B. designed polysome experiments, analysed the data and wrote the manuscript; E.S. performed bioinformatic analysis for the identification of S-AS pairs, designed experiments for the analysis of antisense *Uchl1* expression and analysed the results; S.G. designed the experiments, analysed the results and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.G. (gustinci@sissa.it).

METHODS

Oligonucleotides. The complete list of oligonucleotides used for cloning and for quantitative real-time PCR experiments is included in Supplementary Information (Supplementary Fig. 12).

Plasmids. Full-length DNA sequence of antisense *Uchl1* was amplified via fusion PCR starting from RACE fragment and FANTOM clone 6430596G22 (GenBank AK078321.1) with forward mouse antisense *Uchl1* FL and reverse mouse antisense *Uchl1* FL primers.

Mouse *Uchl1* mRNA was subcloned from FANTOM clone 2900059O22 (GenBank AK013729.1) in the unique PmeI site of pcDNA3.1.

cDNA sequence of human antisense *Uchl1* was amplified from a sample of human brain total RNA (Clontech, 636530) with the primers human 5'F and human 3'R.

Oligonucleotides that target the sequence $-14/+3$ around the TSS of antisense *Uchl1* were annealed and cloned into pSUPERIOR.Neo.GFP vector (OligoEngine) in the BglII/XhoI site. Scrambled sequence was also cloned and used as control.

The antisense *Uchl1* 5' deletion mutant ($\Delta 5'$) was generated by PCR using the oligonucleotides forward mouse antisense *Uchl1*($\Delta 5'$) and reverse mouse antisense *Uchl1* FL. PCR fragment was cloned in the unique EcoRI site in pcDNA3.1.

The antisense *Uchl1* 3' deletion mutant ($\Delta 3'$) was generated by PCR using the forward mouse antisense *Uchl1* FL and reverse mouse antisense *Uchl1*($\Delta 3'$) primers and cloned in the unique EcoRI site in pcDNA3.1.

The antisense *Uchl1*(Δ AS) (Δ Alu + SINEB2) mutant was obtained by subsequent cloning of PCR fragment I (NheI–EcoRI site) and PCR fragment II (EcoRI–HindIII site) into pcDNA3.1. Primers forward mouse antisense *Uchl1* FL NheI and reverse pre-SINE B2 EcoRI were used to generate fragment I; primers forward post-ALU EcoRI and reverse mouse antisense FL HindIII were used for PCR fragment II.

The antisense *Uchl1*(Δ A) (Δ Alu, 1000–1045) mutant was generated with a similar strategy to antisense *Uchl1*(Δ AS). Forward mouse antisense *Uchl1* FL NheI and reverse pre-ALU EcoRI were used for PCR fragment I; forward post-ALU and reverse mouse antisense FL HindIII for PCR fragment II.

The antisense *Uchl1*(Δ S) (Δ SINEB2, 764–934) mutant was obtained with a similar strategy to antisense *Uchl1*(Δ AS). Oligonucleotides forward mouse AS *Uchl1* FL NheI and reverse pre-SINE B2 EcoRI for fragment I; forward post-SINE B2 EcoRI and reverse mouse AS FL HindIII for fragment II.

For antisense *Uchl1*(ASf) (Alu + SINEB2 flipped), PCR fragment obtained with the primers forward SINEB2 inside and reverse Alu flip was cloned in the unique EcoRI site of antisense *Uchl1*(Δ AS).

For antisense *Uchl1*(Sf) (SINEB2 flipped), PCR fragment obtained with forward SINE B2 inside and reverse SINE flip oligonucleotides was cloned in the unique EcoRI site of antisense *Uchl1* Δ SINEB2.

For antisense *Uchl1*(73 bp), the method of 'annealing and primer extension' of two 3'-end overlapping oligonucleotides was used to generate the 73-bp antisense *Uchl1* overlap region. Annealed fragment was obtained with antisense *Uchl1* 73 bp forward and antisense *Uchl1* 73 bp reverse. Fragment was digested with XhoI and EcoRV and ligated into antisense *Uchl1* $\Delta 5'$ plasmid.

The antisense SCR 73-bp mutant was obtained with a similar strategy as antisense *Uchl1*(73 bp). The annealing extension was performed with oligonucleotides with scramble (SCR) sequence (antisense SCR forward and antisense SCR reverse).

Full-length mouse antisense *Uxt* was amplified by PCR starting from FANTOM clone 4833404H03 (GenBank AK029359.1) with specific primers (forward mouse antisense *Uxt* and reverse mouse antisense *Uxt*). PCR fragment was subcloned into pcDNA3.1 using XbaI and HindIII restriction enzymes.

The antisense GFP plasmid was generated with a similar strategy as antisense *Uchl1*(73 bp). Seventy-two base pairs corresponding to nucleotide $-40/+32$ with respect to the ATG of GFP sequence in pEGFP-C2 vector (Clontech) were chosen as target sequence for artificial antisense DNA generation. For annealing, the GFP antisense forward and GFP antisense reverse primers were used.

Cells. MN9D cells were obtained from M. J. Zigmond. Cells were seeded in 100-mm dishes in Dulbecco's modified Eagle's (DMEM) medium containing 10% fetal bovine serum (Invitrogen) supplemented with penicillin (50 units ml^{-1}) and streptomycin (50 units ml^{-1}). For experiments, cells were plated in poly-L-lysine (P2636, Sigma) coated dishes and grown overnight. Approximately 50% confluent cells were treated with $1 \mu\text{M}$ rapamycin (R0395, Sigma) or DMSO vehicle for 45 min.

For the establishment of stable cell lines (shRNA -15/+4, shRNA scrambled, pcDNA 3.1- and AS *Uchl1*(Δ SINEB2)), MN9D cells were seeded in 100-mm Petri dishes and transfected with Lipofectamine 2000 (Invitrogen) according to the manufacturer's instruction. Stable clones were selected by $500 \mu\text{M}$ neomycin (N1142, Sigma). HEK cells (Sigma) were cultured under standard condition in DMEM containing 10% fetal bovine serum supplemented with antibiotics. Transient transfections were done with Lipofectamine 2000 (Invitrogen). For all experiments, S and AS plasmids were transfected at 1:6 ratio.

Animal handling. All animal experiments were performed in accordance with European guidelines for animal care and following SISSA Ethical Committee permissions. Mice were housed and bred in SISSA/CBM non-SPF animal facility, with 12 h dark/light cycles and controlled temperature and humidity. Mice had *ad libitum* access to food and water. C57BL/6 male mice ($n = 5$), 8–10 weeks old, were used for *in situ* hybridization experiments. Laser capture microdissection (LCM) of dopaminergic neurons was performed on 8–10-week-old male TH-GFP/21-31 mice ($n = 3$). Intra-cardiac perfusions were done under total anaesthesia.

RACE and multiplex RT-PCR. The 5' UTR of antisense *Uchl1* was amplified by RACE PCR (GeneRacer, Invitrogen) by MN9D total RNA and cloned into pGEM-T Easy vector (Promega).

qRT-PCR. Total RNA was extracted from cells and mouse tissue samples (adrenal gland, cerebellum, cortex, olfactory bulb, striatum, ventral midbrain, heart, colon, kidney, lung, lymph node, ovary, prostate, spleen, testis, uterus) using Trizol reagent (Invitrogen) according to the manufacturer's instruction. An RNA panel of 20 different normal human tissues (pools consist of at least three tissue donors with full documentation on age, sex, race, cause of death) was obtained from Ambion (AM6000). All RNA samples were subjected to DNase I treatment (Ambion). A total of $1 \mu\text{g}$ of RNA was subjected to retrotranscription using iScript cDNA synthesis kit (BioRad) and Real Time qRT-PCR was carried out using SYBR green fluorescence dye ($2\times$ iQ5 SYBR Green supermix, BioRad). TATA-binding protein (TBP) and RNA polymerase II (RPII) were used as house-keeping genes to normalize different mouse and human tissues as tested by the GeNorm program, version 3.5 (<http://medgen.ugent.be/genorm/>)²⁷. GAPDH and β -actin were used as normalizing controls in all the other qRT-PCR experiments. The amplified transcripts were quantified using the comparative Ct method and the differences in gene expression were presented as normalized fold expression ($\Delta\Delta\text{Ct}$). All of the experiments were performed in duplicate. A heat map graphical representation of rescaled normalized fold expression ($\Delta\Delta\text{Ct}/\Delta\Delta\text{Ct}_{\text{max}}$) was obtained by using Matrix2png (<http://www.bioinformatics.ubc.ca/matrix2png/>). A list of oligonucleotides used for qRT-PCR experiments is in Supplementary Fig. 12.

LCM technology. For LCM, regions of midbrain from TH-GFP/21-31 mice were dissected and incubated in $1\times$ Zincfix solution for 6 h. They were then cryoprotected in 30% sucrose solution at 4°C overnight, embedded in Neg-50 section medium, snap-frozen and left to equilibrate in a cryostat chamber at -21°C for 1 h before sectioning, as described earlier²⁸. Cryostat $14 \mu\text{m}$ midbrain coronal sections were thaw-mounted on Superfrost plus glass slides (Menzel-Glasser) and dopaminergic GFP⁺ cells were gathered via LCM and collected in microfuge (PALM adhesive caps). RNA was immediately extracted using the Absolutely RNA Nanoprep kit (Stratagene), eluted in RNase/DNase free water (Ambion) and retro-transcribed.

Two-colour *in situ* hybridization. After perfusion with 4% formaldehyde, mouse brain was cryoprotected overnight in 30% sucrose. *In situ* hybridization was performed on cryostat slices ($16 \mu\text{m}$). Sense and antisense probes were generated by *in vitro* transcription from the cDNA encoding the distal 600 bp of mouse *Uchl1* cDNA and the last 1,000 bp of mouse antisense *Uchl1*. The probes for *Uchl1* and antisense *Uchl1* were labelled with digoxigenin (DIG labelling, Roche) and biotin (BIO-labelling mix, Roche), respectively. Incorporation of biotin and digoxigenin was checked via a northern blot assay. *In situ* hybridization was performed as described previously²⁹. Slices were pre-treated with 3% hydrogen peroxide for 30 min. Hybridization was performed with probes at a concentration of $1 \mu\text{g ml}^{-1}$ (*Uchl1*) and $3 \mu\text{g ml}^{-1}$ (antisense *Uchl1*) at 60°C for 16 h. For biotinylated RNA detection, streptavidin-HRP (Amersham Bioscience) was used (1:250) for 2 h in TNB buffer (Tris HCl pH 7.5 100 mM, NaCl 150 mM, 0.5% blocking reagent), and signals were visualized using the TSA Cy3 system (Perkin Elmer) after washing in TNT buffer (Tris HCl pH 7.5 100 mM, NaCl 150 mM, 0.05% Tween-20). *In situ* hybridization on DIG-labelled probe was performed with monoclonal anti-DIG antibody after TSA reaction. To combine RNA *in situ* hybridization with immunofluorescence, slices were incubated with anti-tyrosine hydroxylase (TH) antibody 1:1,000 (Chemicon). Signals were then detected with fluorescent dye-conjugated secondary antibody goat anti-rabbit 405 and goat anti-mouse 488. Sections were then washed, mounted with Vectashield (Vector lab) mounting medium and observed with a confocal microscope (Leica).

Western blot. Cells were lysed in SDS sample buffer $2\times$. Proteins were separated in 15% SDS–polyacrylamide gel and transferred to nitrocellulose membranes. Immunoblotting was performed with the following primary antibodies: anti-UCHL1 (3524 Cell Signaling), anti-UXT (11047-1-AP Proteintech Group), anti-p53 (1C12) monoclonal antibody (2524, Cell Signaling) and anti- β -actin (A5441, Sigma). For the mTOR pathway: anti-phospho-p70 S6 kinase (Thr 389) (9234), anti-phospho-4E-BP1 (Ser 65) (9451), anti-p70 S6 kinase (9202), anti-4E-BP1 (9452), anti-phospho-Akt (Ser 473) (3787) were all purchased from Cell Signaling. Signals were revealed after incubation with recommended secondary antibodies

conjugated with horseradish peroxidase by using enhanced chemiluminescence for UCHL1 (WBKLS0500 Immobilized Western Chemiluminescent HRP substrate) and ECL detection reagent (RPN2105, GE Healthcare).

Protein stability. MN9D cells were seeded in 12-well plates overnight and then exposed to $100 \mu\text{g ml}^{-1}$ protein synthesis inhibitor cycloheximide (CHX) for 15 min and rapamycin $1 \mu\text{M}$ or DMSO vehicle control for the following 45 min.

Cellular fractionation. Nucleo-cytoplasmic fractionation was performed using Nucleo-Cytoplasmic separation kit (Norgen) according to the manufacturer's instruction. RNA was eluted and treated with DNase I. The purity of the cytoplasmic fraction was confirmed by qRT-PCR on pre-ribosomal RNA.

Polysome profiles. Polysome profiles were obtained using sucrose density gradients. MN9D cells were treated with $1 \mu\text{g ml}^{-1}$ rapamycin for 35 min, then with $100 \mu\text{g ml}^{-1}$ cycloheximide for 10 min prior to lysis in $150 \mu\text{l}$ lysis buffer (50 mM Tris-HCl pH 7.5, 100 mM NaCl, 30 mM MgCl_2 , $100 \mu\text{g ml}^{-1}$ cycloheximide, 0.1% NP-40, 40 U ml^{-1} RNasin, protease inhibitors cocktail). Whole-cell extracts were clarified at 4°C for 10 min at $15,000g$. The equivalent of 5–10 absorbance units at 254 nm of the clarified cell extract was layered onto 15–55% (w/v) sucrose gradient (50 mM Tris/acetate pH 7.5, 50 mM NH_4Cl , 12 mM MgCl_2 and 1 mM DTT) and centrifuged for 3 h 30 min at $39,000 \text{ r.p.m.}$ in a Beckman SW41Ti rotor at 4°C . The gradient was pumped out by upward displacement and absorbance at 254 nm was monitored using BioLogic LP software (Bio-Rad). One-millilitre fractions were collected, 1 ml Trizol reagent (Invitrogen) was added and RNA was extracted following the manufacturer's instructions. A fixed volume of each RNA sample was then retro-transcribed and the percentage of mRNA in each fraction was calculated as relative Ct value to total RNA.

Metabolic labelling. MN9D were used for analysis of translational rate. Cells were seeded at sub-confluency in 6-well plates, and rapamycin ($1 \mu\text{M}$) or DMSO stimulations were performed for 45 min. Cells were labelled with $5 \mu\text{Ci ml}^{-1}$ of [^{35}S]methionine (Amersham Pharmacia Biotech) for 45 min. Cells were lysed in standard lysis buffer (Tris-HCl 20 mM, NaCl 20 mM, 0.5% Triton X-100) and centrifuged. Supernatants were trichloroacetic-acid-precipitated and filtered on glass fibre discs under vacuum. Discs were counted with scintillation fluid in a β -counter. Total proteins were measured with the standard BCA method. Rate of incorporation was expressed as CPM/total protein ratio (mean \pm s.d.). Experiments were done in triplicate.

Pulse labelling and immunoprecipitation. To monitor *de novo* protein synthesis, HEK cells were transiently transfected with pEGFP plasmid (Clontech) in combination with antisense GFP or empty control vector. After 24 h, medium was replaced with methionine/cysteine-free DMEM for 1 h. Then, cells were labelled with $100 \mu\text{Ci ml}^{-1}$ of [^{35}S]methionine/cysteine (EasyTag, Perkin-Elmer) for 1 h. Labelled cells were collected, lysed in RIPA buffer (150 mM NaCl, 50 mM Tris pH 8, 1 mM EDTA, 1% NP40, 0.5% deoxycholic acid and 0.1% SDS) and used for immunoprecipitation with anti-GFP antibody (Invitrogen) overnight. Immune complexes were isolated with protein G-sepharose beads (Amersham) and separated on 10% SDS-PAGE. Newly translated GFP was visualized by autoradiography. Densitometric analysis was performed on high-resolution images with Photoshop-CS5. Normalization was obtained relative to input.

Northern blot. Polysome fractions were digested with $100 \mu\text{g ml}^{-1}$ proteinase K in 1% SDS and $10 \mu\text{g}$ glycogen (Invitrogen) at 37°C for 1 h. RNA was obtained by phenol/chloroform extraction and re-suspended in formaldehyde/formamide MOPS buffer. Samples were incubated for 5 min at 65°C before being loaded into formaldehyde 1% agarose gel and run at 90 V for 4 h at 4°C . RNA was transferred onto Amersham Hybond-XL nylon membranes and UV-crosslinked. A radio-labelled *Uchl1*-specific complementary RNA probe was transcribed from the same plasmid used for *in situ* hybridization, performing the reaction in the presence of $50 \mu\text{Ci}$ of α - ^{32}P -UTP (Perkin-Elmer). After treatment with DNase I (Ambion), the labelled riboprobe was purified on RNeasy columns (Qiagen). Membranes were pre-hybridized for 3 h at 65°C with NorthernMax prehybridization/hybridization buffer (Ambion) supplemented with $50 \mu\text{g ml}^{-1}$ salmon sperm DNA (Invitrogen), and hybridized with *Uchl1* riboprobe overnight at 65°C in the same buffer. After extensive washes (most stringent conditions were $0.2\times \text{SSC}/0.1\%$ SDS at 65°C), membranes were exposed to autoradiography at -80°C with intensifying screens.

Bioinformatic analysis. For the identification of additional translational activator candidates, we searched for FANTOM3 full-length cDNAs that were non-coding RNAs and overlap the 5' end of coding transcripts in a head-to-head configuration. The filtered set of 8,535 FANTOM3 ncRNA transcripts described previously³⁰ was used as our starting point. Genomic locations of these ncRNA transcripts and RefSeq³¹ coding transcripts were extracted from the alignments in the UCSC Genome browser³² to identify a set of 788 coding-sense–non-coding-antisense pairs. ncRNAs were then checked by RepeatMasker to identify SINEB2-related sequences (<http://www.repeatmasker.org>). This analysis reduced the number of pairs to 127 protein-coding transcripts with overlap at the 5' end (60 with a sense strand version of the repeat, 53 with an antisense version and 14 with both sense and antisense versions).

Alignment of the SINEB2-related elements was then carried out using Clustalw (<http://www.ebi.ac.uk/Tools/clustalw2/index.html>). From this analysis the antisense overlapping transcripts with a repeat most similar to the one of antisense *Uchl1* as well as in the same orientation were chosen for experimental testing (antisense *Uxt*).

Statistical analysis. Statistical analyses were performed with paired two-tailed Student's *t*-test. Results are mean ($n \geq 3$) \pm standard deviation (s.d.).

27. Vandesompele, J. *et al.* Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* **3**, RESEARCH0034 (2002).
28. Biagioli, M. *et al.* Unexpected expression of α - and β -globin in mesencephalic dopaminergic neurons and glial cells. *Proc. Natl Acad. Sci. USA* **106**, 15454–15459 (2009).
29. Ishii, T., Omura, M. & Mombaerts, P. Protocols for two- and three-color fluorescent RNA *in situ* hybridization of the main and accessory olfactory epithelia in mouse. *J. Neurocytol.* **33**, 657–669 (2004).
30. Nordström, K. J. *et al.* Critical evaluation of the FANTOM3 non-coding RNA transcripts. *Genomics* **94**, 169–176 (2009).
31. Maglott, D. R., Katz, K. S., Sicotte, H. & Pruitt, K. D. NCBI's LocusLink and RefSeq. *Nucleic Acids Res.* **28**, 126–128 (2000).
32. Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).

Serine is a natural ligand and allosteric activator of pyruvate kinase M2

Barbara Chaneton^{1*}, Petra Hillmann^{2*}, Liang Zheng¹, Agnès C. L. Martin², Oliver D. K. Maddocks¹, Achuthanunni Chokkathukalam³, Joseph E. Coyle², Andris Jankevics^{3,4}, Finn P. Holding², Karen H. Vousden¹, Christian Frezza^{1†}, Marc O'Reilly² & Eyal Gottlieb¹

Cancer cells exhibit several unique metabolic phenotypes that are critical for cell growth and proliferation¹. Specifically, they over-express the M2 isoform of the tightly regulated enzyme pyruvate kinase (PKM2), which controls glycolytic flux, and are highly dependent on *de novo* biosynthesis of serine and glycine². Here we describe a new rheostat-like mechanistic relationship between PKM2 activity and serine biosynthesis. We show that serine can bind to and activate human PKM2, and that PKM2 activity in cells is reduced in response to serine deprivation. This reduction in PKM2 activity shifts cells to a fuel-efficient mode in which more pyruvate is diverted to the mitochondria and more glucose-derived carbon is channelled into serine biosynthesis to support cell proliferation.

Metabolic fluxes in cancer cells are different from those in non-transformed cells¹. In particular, a shift from oxidative phosphorylation to aerobic glycolysis has been demonstrated, which is promoted by the M2 isoform of pyruvate kinase³. PKM2 catalyses the final step of glycolysis, converting phosphoenolpyruvate (PEP) to pyruvate (Supplementary Fig. 1). Interestingly, PKM2, which is the predominant isoform in cancer cells^{4,5}, has low basal enzymatic activity compared to the constitutively active splice variant PKM1 (ref. 6). Another metabolic pathway recently demonstrated to be crucial for cancer cell survival is the serine-biosynthesis pathway^{7–9}. We investigated a potential mechanistic link between the two pathways in cancer cells, whereby a reduction in overall pyruvate kinase activity through the preferential expression of PKM2 would cause the build-up of glycolytic intermediates for channelling into the serine-biosynthetic pathway. To test this hypothesis, we used human colon carcinoma HCT116 cells, which predominantly express the PKM2 isoform (Fig. 1a and Supplementary Fig. 2). Two discrete short hairpin RNA (shRNA) pools were used to generate two independent HCT116-derived cell lines, shPKMa and shPKMb, in which the expression of both the PKM1 and PKM2 isoforms was simultaneously and stably silenced (Fig. 1a and Supplementary Fig. 2c). Despite achieving a greater than 90% reduction in PKM1 and PKM2 (hereafter termed PKM for simplicity) messenger RNA and protein levels, compared to cells expressing non-targeting shRNA (shCtrl), no compensatory transcriptional induction of the PKL or PKR isoforms was observed in the shPKM cells (Supplementary Fig. 2). In line with this, liquid chromatography–mass spectrometry (LC–MS) analysis of the steady-state levels of metabolites revealed a 100-fold increase in PEP concentration in shPKM cells accompanied by a ~50% decrease in pyruvate levels, demonstrating a reduction in intracellular pyruvate kinase activity (Fig. 1b). The stable silencing of PKM in HCT116 cells did not alter cell proliferation rates or steady-state levels of ATP (Fig. 1c, d). By contrast, the proliferation rates of HT29 and SW620 colon cancer cells were more sensitive to PKM silencing (Supplementary Fig. 3a). Regardless of the effect on cellular proliferation rates, PKM silencing universally increased the oxygen-consumption

rates by ~30%, with a corresponding decrease in extracellular-acidification rates, indicators of increased oxidative phosphorylation and decreased glycolysis, respectively (Supplementary Fig. 3b, c). Because pyruvate kinase catalyses an important ATP-producing step in glycolysis, the stability of intracellular ATP levels could be explained by this compensatory increase in oxidative phosphorylation in response to PKM silencing. Thus, despite the predominant expression of PKM2 in HCT116 cells, these cells still exhibit sufficient pyruvate kinase activity to convert PEP to pyruvate and to facilitate aerobic glycolysis.

Whereas PKM silencing caused a large increase in PEP concentration, pyruvate levels were decreased to a lesser extent (Fig. 1b). There are several possible explanations for this. First, residual PKM could still generate pyruvate, albeit at a lower rate. Second, pyruvate can be synthesized from carbon sources other than glucose. Finally, pyruvate can also be generated from PEP through a pyruvate-kinase-independent mechanism¹⁰, although this alternative pathway was not increased in the knockdown cells (Supplementary Fig. 4).

In order to study the fate of glucose in PKM-inhibited cells, shCtrl and shPKM cells were incubated in media containing uniformly ¹³C-labelled glucose (U-¹³C-glucose) and cells were extracted at different time points. Several glucose-derived metabolites were tracked by LC–MS (Fig. 2 and Supplementary Fig. 5), including pyruvate and PEP. The ratio between these two metabolites at an early time point after glucose labelling was validated as a reliable measure of PKM2 activity using an activator of PKM2 (Supplementary Figs 5 and 6 and Supplementary Discussion).

In the cytosol, pyruvate is metabolized to lactate by lactate dehydrogenase (LDH) and the resulting lactate contains three glucose-derived carbons. In addition, pyruvate is translocated to the mitochondria where it is oxidized and de-carboxylated to acetyl-CoA, which enters the tricarboxylic acid (TCA) cycle to form citrate, contributing two carbon atoms from glucose. When cells were incubated with U-¹³C-glucose, both glucose-derived lactate and citrate were detected by LC–MS (¹³C₃-lactate and ¹³C₂-citrate). Blocking PKM activity shifted the metabolism of glucose away from lactate production in the cytosol to citrate production in the mitochondria (Fig. 2 and Supplementary Fig. 5). Heavier isotopomers of citrate (particularly ¹³C₄-citrate) were detected in cells incubated for a longer time period (4 h) with U-¹³C-glucose, an indication of further oxidation and generation of citrate in the TCA cycle (Supplementary Fig. 7). These results are in line with the observed increase in oxygen consumption of shPKM cells (Supplementary Fig. 3).

An increased metabolic flux into the serine and glycine biosynthetic pathway was also observed in cells with reduced pyruvate kinase activity, as determined by the accumulation of glucose-derived ¹³C in these amino acids (Fig. 2 and Supplementary Fig. 5). This is the first

¹Cancer Research UK, The Beatson Institute for Cancer Research, Switchback Road, Glasgow G61 1BD, Scotland, UK. ²Astex Pharmaceuticals, 436 Cambridge Science Park, Milton Road, Cambridge CB4 0QA, UK. ³Institute of Molecular, Cell and Systems Biology, College of Medical, Veterinary and Life Sciences, Joseph Black Building, B3.09, University of Glasgow, Glasgow G12 8QQ, Scotland, UK. ⁴Groningen Bioinformatics Centre, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Groningen 9747 AG, The Netherlands. [†]Present address: MRC Cancer Cell Unit, Hutchison/MRC Research Centre, Hills Road, Cambridge, CB2 0XZ, UK.

*These authors contributed equally to this work.

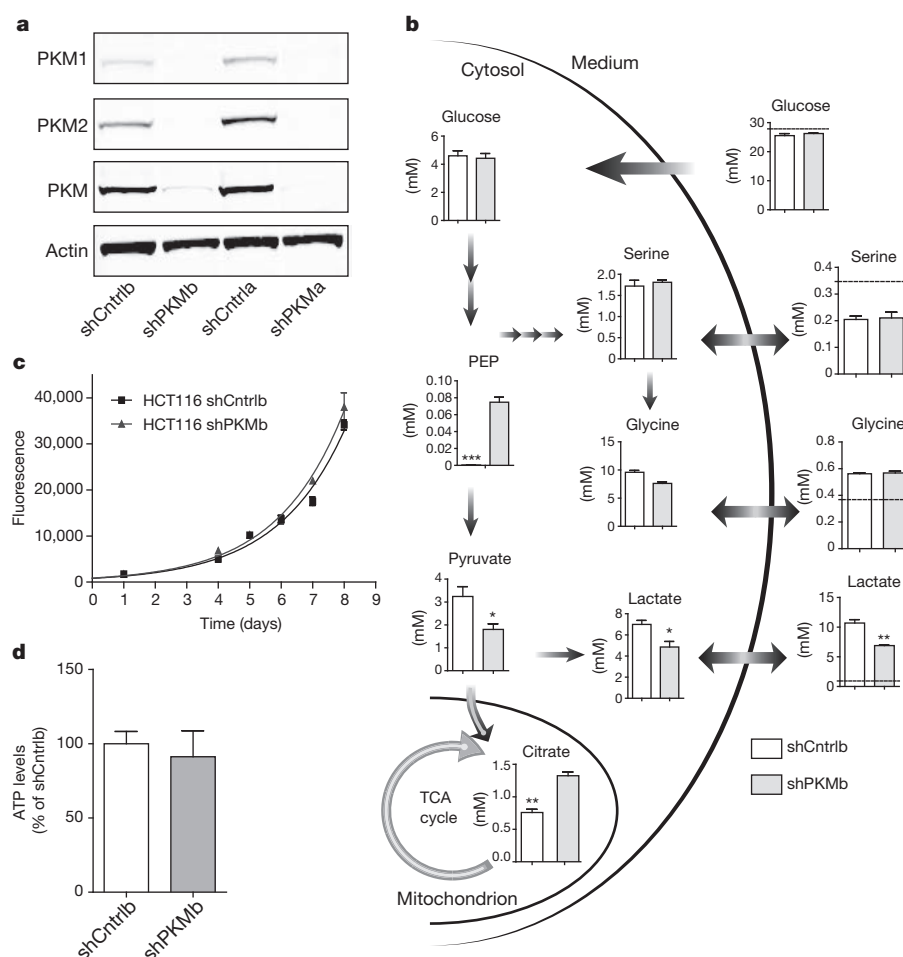


Figure 1 | Characterization of PKM-silenced HCT116 cells. **a**, PKM1 and PKM2 protein levels were detected by western blot. Actin was used as a loading control. **b**, Quantified intracellular metabolite concentrations and the uptake or secretion of extracellular metabolites in control and shPKM HCT116 cells. For extracellular metabolites, the dashed line indicates the initial levels in the

medium and the graph represents the levels after 24 h incubation. **c**, The proliferation rate of the indicated cell lines. **d**, Intracellular ATP levels normalized to protein concentration in the cell extracts. Data are mean \pm s.e.m. of triplicates and are representative of three (two in **b**) independent experiments. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

direct evidence that low pyruvate kinase activity can drive serine and glycine biosynthesis, and demonstrates an important link between key metabolic processes observed in cancer, namely preferential PKM2 expression, aerobic glycolysis and serine biosynthesis. The relative contribution of glucose-derived carbons to serine and glycine is low (Supplementary Fig. 7), which is attributable to the presence of unlabelled serine and glycine in the growth media of the cells. Indeed, when cells were incubated for 12 h in serine- and glycine-free media, more than 80% of the intracellular serine and glycine was glucose derived (Supplementary Fig. 8a). Nevertheless, *de novo* synthesis of serine and glycine in cells starved of these two amino acids was insufficient to recover their steady-state levels (Fig. 3a), indicating a rapid use of newly synthesized amino acids. Importantly, the absence of extracellular serine and glycine had a pronounced inhibitory effect on pyruvate kinase activity, as demonstrated by a 100% increase in PEP and a 30% decrease in pyruvate (Fig. 3a). Akin to PKM silencing, serine deprivation reduced cytosolic lactate production and increased mitochondrial citrate production by $\sim 60\%$ (Fig. 3a). Moreover, cellular deprivation of serine and glycine for 12 h followed by 30-min incubation with U- ^{13}C -glucose resulted in a 50% decrease in the labelled pyruvate/PEP ratio (Fig. 3b). These data indicate that serine and glycine deprivation decreases PKM2 activity in cells, such that more glucose-derived carbon is channelled into serine and glycine biosynthesis.

PKM2 is a tightly regulated enzyme that responds not only to the availability of PEP and ADP substrates, but also to the upstream

glycolytic metabolite fructose-1,6-bisphosphate (FBP) and to phosphorylation events^{2,11–14}. Previous studies have proposed that pyruvate kinase activity may be regulated by amino acids^{15–17}. Therefore, the ability of serine or glycine to stimulate PKM2 in cells was tested. Serine hydroxymethyltransferase converts serine to glycine and vice versa, and therefore cells were starved of both amino acids overnight before 30-min incubation with either serine or glycine. The cells incubated with serine or glycine showed an increase in the intracellular levels of the added amino acid only, and the levels of the other amino acid were unaffected (Supplementary Fig. 8b). When serine- and glycine-starved cells were incubated for 30 min with serine together with U- ^{13}C -glucose, intracellular pyruvate kinase activity was increased relative to the starved cells. However, glycine did not stimulate intracellular pyruvate kinase activity (Supplementary Fig. 8c).

In order to substantiate the results observed in cells, recombinant human PKM2 activity was further analysed *in vitro*. Serine was demonstrated to activate recombinant PKM2 with a half-maximal activation concentration (AC_{50}) of 1.3 mM (Fig. 4a), a level that is within the physiological range of intracellular serine concentrations (Fig. 1b). Isothermal titration calorimetry (ITC) was used to determine the dissociation constant (K_d) of the PKM2–serine interaction as 0.20 mM (Fig. 4b) and the titration curve was consistent with a 1:1 PKM2–monomer:serine ratio. These results not only demonstrate direct interactions between serine and PKM2; they also suggest that the serine concentration required for such interactions is well within

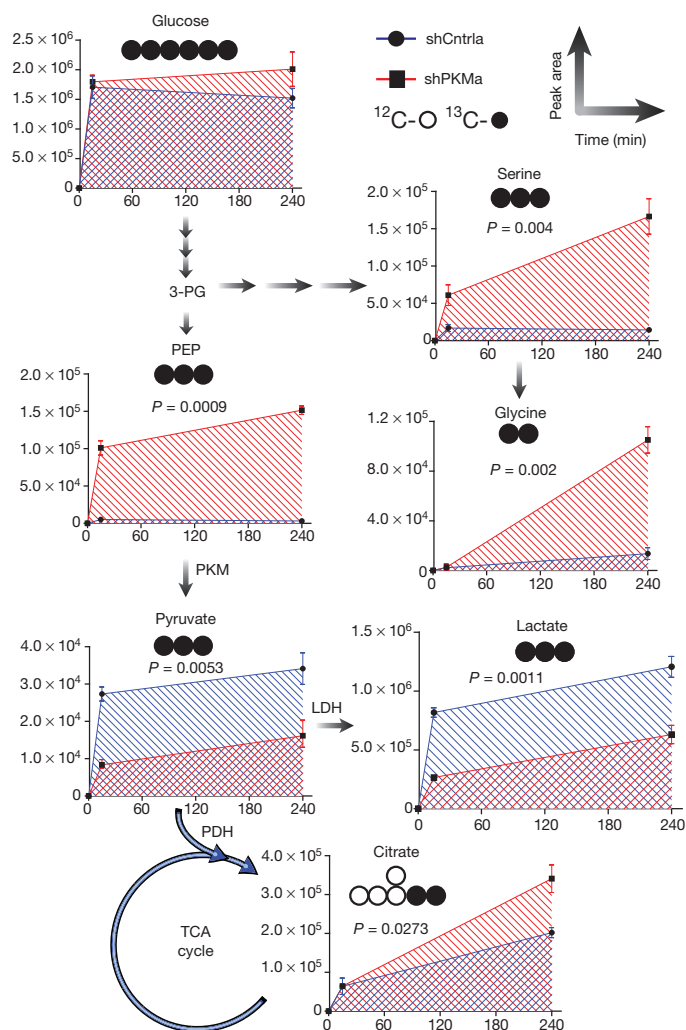


Figure 2 | The effect of PKM silencing on glycolytic flux. Cells were incubated with U- ^{13}C -glucose and the abundance of the main glucose-derived metabolite isotopomers was analysed at the indicated time points. The cumulative intensities of each labelled metabolite are presented in blue (shCntrl) or red (shPKM). The white and black circles under each metabolite illustrate ^{12}C and ^{13}C labelling, respectively. All metabolic quantifications are presented as mean \pm s.e.m. of triplicates and are representative of three independent experiments. 3-PG, 3-phosphoglycerate; PDH, pyruvate dehydrogenase. *P* values at 4 h are shown where significant.

the physiological range of intracellular serine levels. In agreement with the results observed in cells, glycine could not directly activate PKM2 *in vitro* (Fig. 4a). Similarly to FBP, serine lowered the Michaelis constant (K_m) of PKM2 for PEP by 2.3-fold, effectively increasing the affinity of PKM2 for this substrate (Fig. 4c). Serine was found to be the only standard amino acid that could activate PKM2 (Supplementary Fig. 9a). A fragment-based crystallographic screen¹⁸ versus human PKM2 revealed the amino acids L-alanine, L-cysteine, L-threonine and L-serine bound to a previously uncharacterized binding pocket on PKM2. Crystallographic soaking experiments revealed L-serine bound to PKM2, with a single L-serine molecule bound to each of the monomers comprising the PKM2 tetramer (Fig. 4d and Supplementary Tables 1 and 2). L-serine made multiple hydrogen-bonding interactions with PKM2 (see Supplementary Fig. 9 and Supplementary Discussion). Additional interactions afforded by the serine side-chain hydroxyl group rationalize the low affinity of glycine for PKM2 and may contribute to the unique ability of serine to activate PKM2 (ref. 16). The amino acid binding pocket was vacant in crystals not soaked with serine (Supplementary Fig. 9b).

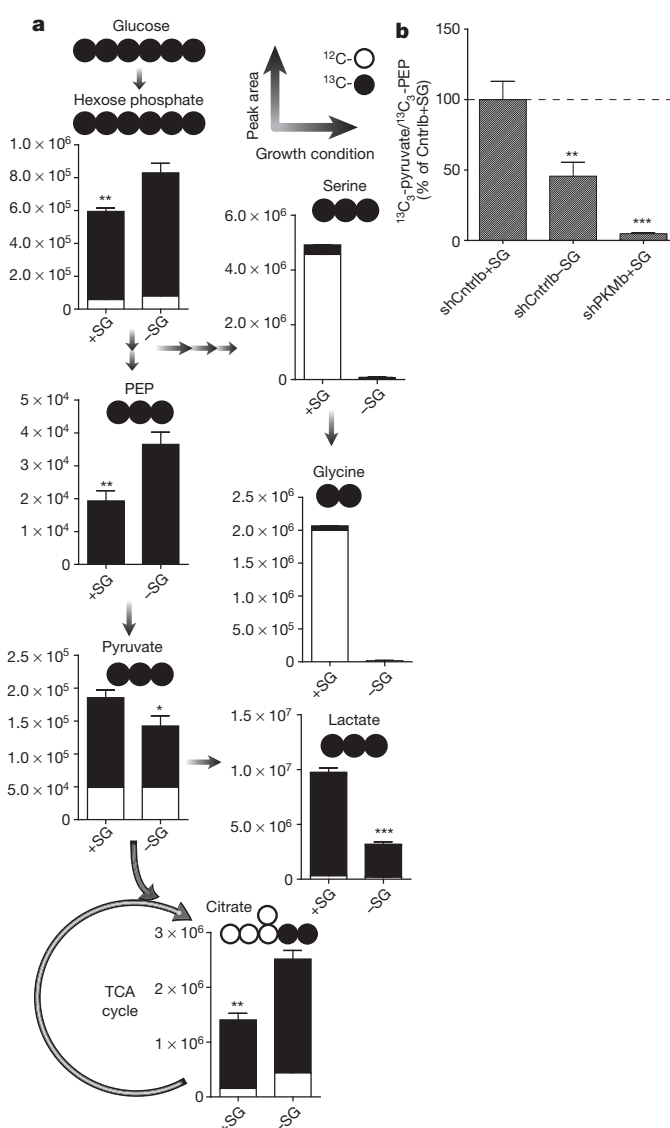


Figure 3 | Serine and glycine deprivation changes glucose metabolism.

a, HCT116 cells were incubated for 12 h with U- ^{13}C -glucose in the presence (+SG) or absence (−SG) of serine and glycine. The abundance of the main glucose-derived isotopomers (black bars) and the unlabelled fraction (white bars) of the indicated metabolites were analysed. The circles represent the relevant isotopomers as in Fig. 2. **b**, Intracellular pyruvate kinase activity is represented as the ratio between glucose-derived ($^{13}\text{C}_3$ -) pyruvate and PEP, 30 min after labelling with U- ^{13}C -glucose. Results were normalized to shCntrl + SG. All results are presented as mean \pm s.e.m. of triplicates and are representative of two independent experiments. **P* < 0.05, ***P* < 0.01, ****P* < 0.001.

On the basis of the above observations, a recombinant PKM2(H464A) mutant was produced and demonstrated neither direct binding to, nor activation by, serine (Fig. 4b, e). These data confirm that the amino-acid-binding site identified in the soaking experiments is the only amino-acid-binding site on the PKM2 protein. Despite its inability to bind serine, the PKM2(H464A) mutant was activated by FBP in a similar way to wild-type PKM2. Conversely, a S437Y mutant of PKM2, which cannot bind FBP¹⁹, was activated by serine (Fig. 4e, f). These results demonstrate that PKM2 is independently activated by either FBP or serine, and that both molecules could contribute to PKM2 regulation in response to glucose and/or amino acid deprivation *in vivo*. PKM2 regulation by serine in tumours is important as the serine concentration in the blood is ~ 20 -fold below that of glucose, hence under an interrupted blood supply, a considerable deficit in serine supply

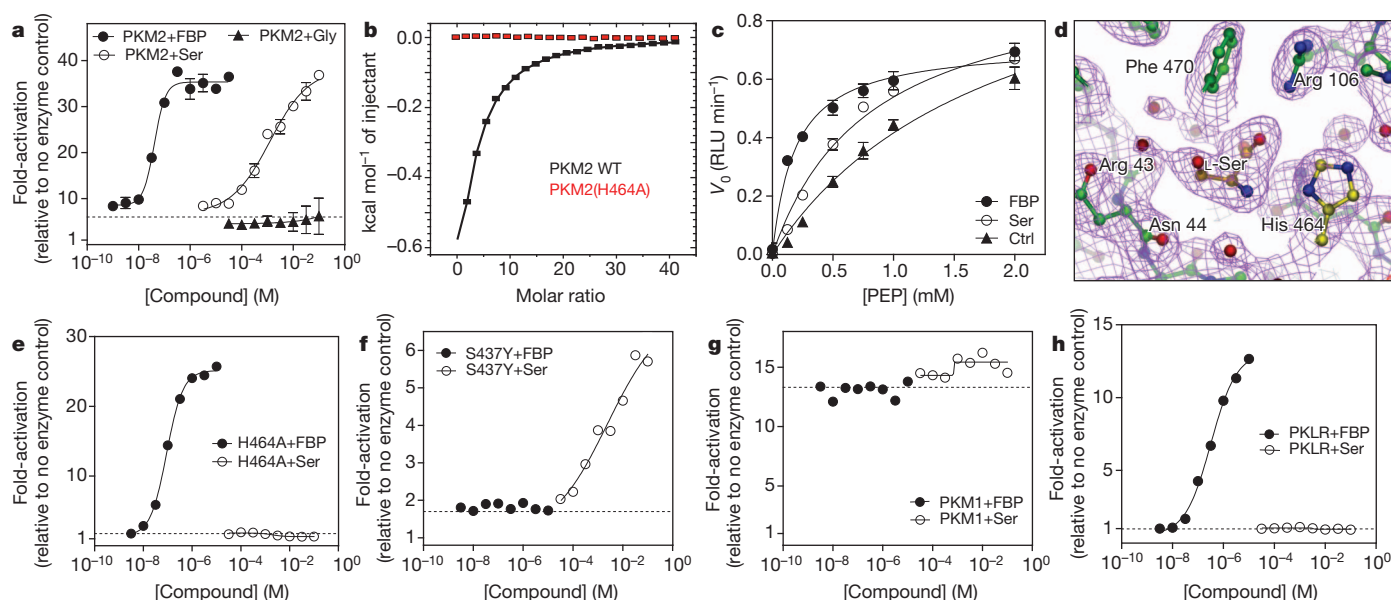


Figure 4 | Serine is an allosteric activator of PKM2. **a**, *In vitro* activity of recombinant human PKM2 was analysed in the presence of increasing concentrations of FBP (●), serine (○) or glycine (▲). **b**, Serine binding to PKM2 wild-type (WT) or PKM2(H464A) mutant was measured by ITC. The K_d of serine for wild-type PKM2 was measured as 200 μ M. No serine binding to PKM2(H464A) was detected. **c**, The initial PKM2 reaction rate (V_0 (relative luminescence units (RLU) per min)) was measured at different PEP concentrations in the presence of 50 μ M FBP (●), 100 mM serine (○) or vehicle (▲). K_m values for PEP were determined as 1.9 mM, 0.81 mM and 0.19 mM in the presence of vehicle, serine and FBP, respectively. For **a** and **c**, results are presented as mean \pm s.e.m. of three experiments. **d**, 2.3 Å $2F_o - F_c$

map (purple) contoured at 1σ for the final, refined structure of L-serine (orange) bound to PKM2 (green). The side chain of His 464, which was subsequently mutated to alanine, is shown in yellow. **e**, **f**, *In vitro* activity of PKM2 mutants H464A (**e**) and S437Y (**f**) was analysed in the presence of increasing concentrations of FBP (●) or serine (○). **g**, **h**, *In vitro* activity of PKM1 (**g**) and PKLR (**h**) was analysed in the presence of increasing concentrations of FBP (●) or serine (○). For **a** and **e–h**, the signal was normalized to controls containing no enzyme. The basal activity of the studied pyruvate kinase in each panel in the presence of vehicle control is indicated by the dotted line. For **e–h**, data are presented as the mean of duplicate determinations and are representative of three independent experiments.

would occur. Serine is crucial for multiple metabolic pathways required for cell growth and proliferation, including phospholipid, purine and glutathione biosynthesis, as well as being a methyl source for single-carbon metabolism. In addition, serine levels are depleted at the periphery of solid tumours, hence *de novo* serine biosynthesis is critical for tumour growth^{20,21}.

The allosteric regulation of pyruvate kinase is associated with complex structural changes^{11,14,22}. Phosphorylation of Tyr 105, adjacent to Arg 106 in the serine-binding pocket, has been shown to modulate PKM2 activity and FBP binding¹², and oxidation of Cys 358, which is proximal to the serine-binding pocket, inhibits PKM2 (ref. 23) (Supplementary Fig. 9e). This indicates that the PKM2 serine-binding pocket, and proximal residues, constitute a key structural regulatory node for PKM2.

The predominant isoform of pyruvate kinase in cancer cells is PKM2; however, the serine-binding site identified here is conserved in PKM1, PKL and PKR. Therefore, the possibility that serine may be a universal pyruvate kinase regulator was tested *in vitro*. PKM1 demonstrated a high degree of basal activity in the absence of exogenous activators, and was refractory to both FBP and serine activation (Fig. 4g). By contrast, a recombinant protein corresponding to the common portion of PKL and PKR (referred to as PKLR) was completely inactive in its basal state and was robustly activated by FBP. Notably, PKLR was not activated by serine (Fig. 4h). These results indicate that only PKM2-expressing cells can respond to changes in serine availability and support shuttling of glucose-derived carbon into serine biosynthesis after serine deprivation.

This work provides a new understanding of the relationship between glucose and amino acid metabolism. Serine biosynthesis is an anabolic pathway required for growth and proliferation²⁴. However, it recruits carbon away from the energy-production pathway of glucose utilization. Here, to our knowledge, we present for the first time an elucidation of the mechanism that tightly controls the metabolic bifurcation of glucose-derived carbon. The control of PKM2 activity through serine

availability provides a rheostat-like mechanism. When serine is abundant, PKM2 is fully active, enabling the maximal use of glucose through glycolysis. However, when the steady-state levels of serine drop below a critical point, an immediate attenuation of PKM2 activity occurs. This enables the fast shuttling of glucose-derived carbon to serine biosynthesis, compensating for the serine shortfall and enabling growth and proliferation in the absence of these amino acids (Supplementary Fig. 1). Finally, by activating PKM2, serine supports aerobic glycolysis and lactate production, events that are critical for cancer cell growth and survival.

METHODS SUMMARY

shCtrl and shPKM cells were obtained by infecting cells with control non-targeting shRNA or shPKM lentiviral particles, respectively. mRNA and protein levels of pyruvate kinase isoforms were determined by quantitative PCR and western blot. Cell proliferation was measured using the Alamar Blue assay. ATP levels were measured using a bioluminescent assay kit. Oxygen-consumption rate and extracellular-acidification rate were measured using a Seahorse XF24. The quantification of intracellular and extracellular metabolites was performed by LC-MS. Isotopomer distribution of intracellular metabolites after labelling with U - ^{13}C -glucose was analysed with the PeakML.Isotope.TargettedIsotopes() function of the mzmatch.R library. *In vitro* PKM2 activity was measured by the quantification of ATP production using luminescent Kinase-Glo Plus reagent. ITC experiments were performed on a MicroCal VP-ITC isothermal titration calorimeter. His-tagged human PKM2 was purified, concentrated, crystallized and then soaked in the presence or absence of serine and X-ray diffraction data were collected.

Full Methods and any associated references are available in the online version of the paper.

Received 13 November 2011; accepted 23 August 2012.

Published online 14 October; corrected online 14 November 2012 (see full-text HTML version for details).

1. Tennant, D. A., Duran, R. V. & Gottlieb, E. Targeting metabolic transformation for cancer therapy. *Nature Rev. Cancer* **10**, 267–277 (2010).

2. Chaneton, B. & Gottlieb, E. Rocking cell metabolism: revised functions of the key glycolytic regulator PKM2 in cancer. *Trends Biochem. Sci.* **37**, 309–316 (2012).
3. Christofk, H. R. *et al.* The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth. *Nature* **452**, 230–233 (2008).
4. Altenberg, B. & Greulich, K. O. Genes of glycolysis are ubiquitously overexpressed in 24 cancer classes. *Genomics* **84**, 1014–1020 (2004).
5. Mazurek, S., Boschek, C. B., Hugo, F. & Eigenbrodt, E. Pyruvate kinase type M2 and its role in tumor growth and spreading. *Semin. Cancer Biol.* **15**, 300–308 (2005).
6. Yamada, K. & Noguchi, T. Nutrient and hormonal regulation of pyruvate kinase gene expression. *Biochem. J.* **337**, 1–11 (1999).
7. Possemato, R. *et al.* Functional genomics reveal that the serine synthesis pathway is essential in breast cancer. *Nature* **476**, 346–350 (2011).
8. Locasale, J. W. *et al.* Phosphoglycerate dehydrogenase diverts glycolytic flux and contributes to oncogenesis. *Nature Genet.* **43**, 869–874 (2011).
9. Pollari, S. *et al.* Enhanced serine production by bone metastatic breast cancer cells stimulates osteoclastogenesis. *Breast Cancer Res. Treat.* **125**, 421–430 (2011).
10. Vander Heiden, M. G. *et al.* Evidence for an alternative glycolytic pathway in rapidly proliferating cells. *Science* **329**, 1492–1499 (2010).
11. Dombrauckas, J. D., Santarsiero, B. D. & Mesecar, A. D. Structural basis for tumor pyruvate kinase M2 allosteric regulation and catalysis. *Biochemistry* **44**, 9417–9429 (2005).
12. Hitosugi, T. *et al.* Tyrosine phosphorylation inhibits PKM2 to promote the Warburg effect and tumor growth. *Sci. Signal.* **2**, ra73 (2009).
13. Christofk, H. R., Vander Heiden, M. G., Wu, N., Asara, J. M. & Cantley, L. C. Pyruvate kinase M2 is a phosphotyrosine-binding protein. *Nature* **452**, 181–186 (2008).
14. Ashizawa, K., Willingham, M. C., Liang, C. M. & Cheng, S. Y. *In vivo* regulation of monomer-tetramer conversion of pyruvate kinase subtype M2 by glucose is mediated via fructose 1,6-bisphosphate. *J. Biol. Chem.* **266**, 16842–16846 (1991).
15. Spellman, C. M. & Fottrell, P. F. Similarities between pyruvate kinase from human placenta and tumours. *FEBS Lett.* **37**, 281–284 (1973).
16. Eigenbrodt, E., Leib, S., Kramer, W., Friis, R. R. & Schoner, W. Structural and kinetic differences between the M2 type pyruvate kinases from lung and various tumors. *Biomed. Biochim. Acta* **42**, S278–S282 (1983).
17. Ye, J. *et al.* Pyruvate kinase M2 promotes de novo serine synthesis to sustain mTORC1 activity and cell proliferation. *Proc. Natl Acad. Sci. USA* **109**, 6904–6909 (2012).
18. Davies, T. G. & Tickle, I. J. Fragment screening using X-ray crystallography. *Top. Curr. Chem.* **317**, 33–59 (2012).
19. Allali-Hassani, A. *et al.* A survey of proteins encoded by non-synonymous single nucleotide polymorphisms reveals a significant fraction with altered stability and activity. *Biochem. J.* **424**, 15–26 (2009).
20. Medina, M. A., Marquez, J. & Nunez de Castro, I. Interchange of amino acids between tumor and host. *Biochem. Med. Metab. Biol.* **48**, 1–7 (1992).
21. Márquez, J., Sanchez-Jimenez, F., Medina, M. A., Quesada, A. R. & Nunez de Castro, I. Nitrogen metabolism in tumor bearing mice. *Arch. Biochem. Biophys.* **268**, 667–675 (1989).
22. Mattevi, A., Bolognesi, M. & Valentini, G. The allosteric regulation of pyruvate kinase. *FEBS Lett.* **389**, 15–19 (1996).
23. Anastasiou, D. *et al.* Inhibition of pyruvate kinase M2 by reactive oxygen species contributes to cellular antioxidant responses. *Science* **334**, 1278–1283 (2011).
24. de Koning, T. J. *et al.* L-serine in disease and development. *Biochem. J.* **371**, 653–661 (2003).

Supplementary Information is available in the online version of the paper.

Acknowledgements The work performed at the Beatson Institute for Cancer Research was supported by Cancer Research UK. We thank D. Sumpton for technical support with two-dimensional gel electrophoresis and N. Thompson, N. Wallis and M. Jones for comments provided during manuscript preparation. We would also like to thank D. M. Sabatini for the Scramble shRNA plasmid used as a control (shCntrl) and the Structural Genomics Consortium for providing us with the PKM2 expression plasmid from their collection. We thank A. King for editorial work and S. Tardito for graphical help.

Author Contributions M.O. and E.G. conceived the project and wrote the manuscript with the help of B.C., P.H. and C.F. L.Z., B.C. and C.F. performed the LC–MS assay and analysed the raw data. A.C. and A.J. analysed the LC–MS data and identified the different isotopomers of each metabolite. A.C.L.M. performed the *in vitro* enzymatic activity, J.E.C. performed the ITC, M.O. generated the point mutant constructs, purified the proteins and solved the crystal structure. F.P.H. performed the LC–MS validation of the point mutant constructs. O.D.K.M. and K.H.V. performed, analysed and discussed the long-term serine and glycine starvation experiment. B.C. and P.H. generated and characterized the cell lines and performed all other experiments and data analysis. All the authors discussed the results and commented on the manuscript.

Author Information Atomic coordinates and structure factors for the PKM2 crystal structures have been deposited in the Protein Data Bank (PDB) under accession code 4B2D. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.G. (e.gottlieb@beatson.gla.ac.uk) or M.O. (marc.oreilly@astx.com).

METHODS

Cell culture. HCT116 and HT29 colon cancer cells were maintained at 37 °C and 5% CO₂ in high glucose DMEM (21969-035, Invitrogen) supplemented with 10% FBS and 2 mM L-glutamine. SW620 colon cancer cells were maintained at 37 °C and 5% CO₂ in RPMI (12633-020, Invitrogen) supplemented with 20% FBS and 2 mM L-glutamine. Stable PKM knockdown and control cell lines were cultured in the same media containing additional 2 µg ml⁻¹ puromycin.

Stable PKM silencing. HCT116, SW620 and HT29 cells were infected with lentiviral particles containing control shRNA (shCntrlb) (sc-108080) or a pool of three PKM shRNAs (shPKMb) (sc-62820) (5'-GATCCCTGTGGCTCTAGACACTAATTCAAGAGATTA-GTGTCTAGAGCCACAGTTTT-3'; 5'-GATCCGTCTGGAGAAACAGCCAAATTC-AAGAGATTGGCTGTTTCTCCAGACTTTT-3'; 5'-GATCCGTCTGGAGAAACAGC-CAAATTCAGAGATTGCTGTCTCTCCAGACTTTT-3') (Santa Cruz Biotechnology) according to the manufacturer's instructions. Infected cells were selected using 6 µg ml⁻¹ puromycin and shPKM clones were analysed for PKM1 and PKM2 expression levels using western blot analysis and quantitative PCR (qPCR). A different set of plasmids containing PKM shRNA (shPKMa) was bought from Openbiosystems (TRCN000037610) (5'-CCGG-GAAGGGAAGAACATCAAGAT-CTCGAG-ATCTTGATGTTCTTTCCCTTC-TTTT-3') and (TRCN000037611) (5'-CCGG-CGGGTGAACCTTGGCATGAAT-CTCGAG-ATTCATGGCAAGTTCACCCG-TTTT-3'). Scramble shRNA (shScramble) (5'-CCTAAGGTTAAGTCGCCCTCGCTCGAGCGAGGG-CGACTTAACCTTAGG-3') (Addgene plasmid no. 1864) was used as a control (shCntrla). HCT116 cells were infected with both pLKO.1-shPKM or pLKO.1-shScramble and selected using 2 µg ml⁻¹ puromycin for 2 weeks and PKM-silenced clones were analysed for PKM1 and PKM2 expression levels using western blot analysis and qPCR.

mRNA extraction and qPCR analyses. 4 × 10⁵ cells were plated in a 6-well plate and were lysed after 2 days in RNeasy lysis buffer (Qiagen). Lysates were passed through QIAshredder columns (Qiagen) and mRNA was isolated using the RNeasy kit following the manufacturer's instructions. RNA was quantified and quality-controlled using an Eppendorf biophotometer and Eppendorf single-sealed cuvettes, UVette (Eppendorf UK Limited). For PCR analyses 1 µg mRNA was retro-transcribed into complementary DNA using High Capacity RNA-to-cDNA (AB, Life Technologies Corporation). In brief, 0.5 µM primers, 1X Fast SYBR Green Master mix (AB, Life Technologies Corporation) and 1 µl of a 1:10 dilution of cDNA in a final volume of 20 µl were used. qPCR was performed on the 7500 Fast Real-Time PCR System (Life Technologies Corporation) and expression levels of the indicated genes were calculated using the $\Delta\Delta C_t$ method by the appropriate function of the software using actin as calibrant. The PCR program was: 20 s at 95 °C followed by 40 cycles of 3 s at 95 °C and 30 s at 60 °C. Finally the melting curve was performed, which was used to confirm the presence of single PCR products. Primers are as follows: β -actin-forward primer: 5'-TCCATCATGAAGTGTGACGT-3'; β -actin-reverse primer: 5'-TACTCTGCTTCTGATCCAC-3'; PKM1-forward primer: 5'-GAGGCAGCCATGTTCCAC-3'; PKM1-reverse primer: 5'-TGCCAGACTCCGTCAGAACT-3'; PKM2-forward primer: 5'-CAGAGGCTGCCATCTACCAC-3'; PKM2-reverse primer: 5'-CCAGACTTGGTGAGGACGAT-3'. PKL-forward primer: 5'-CTGGTGATTGTGGTGACAGG-3' PKL-reverse primer: 5'-TGGGCTGGAGAACGTAGACT-3' PKR-forward primer: 5'-CAATTTGGCATTGAAAGTGG-3'; PKR-reverse primer: 5'-CCTGTACCAACATACCAG-3'.

Immunoblotting. 4 × 10⁵ cells were plated in a 6-well plate and were lysed after 2 days in radio-immunoprecipitation assay buffer (RIPA) (150 mM sodium chloride, 1.0% NP-40, 0.5% sodium deoxycholate and 50 mM Tris, pH 8.0) supplemented with a 1:100 dilution of the protein inhibitors cocktail (Sigma). Protein concentration was determined using the bicinchoninic acid assay (Thermoscientific) using BSA as standard (Thermoscientific). Equal amounts of protein were loaded into 12% SDS-PAGE gels and electrophoretically separated using Tris-glycine SDS running buffer. After SDS-PAGE, proteins were transferred onto 0.22 µm nitrocellulose (Millipore) and probed with antibodies, all at 1:1,000 dilution in 5% non-fat milk. PKM1 antibody was custom-made by PolyPeptide Laboratories using the following peptide sequence: CLVRASSHSTDLMEAMAMGS. The PKM2 (cat no. 3198) and PKM (no. 3186) antibodies were purchased from Cell Signaling Technology. The anti-actin antibody (mouse monoclonal AC-40) was purchased from Sigma. For detection, membranes were incubated with either donkey-anti rabbit (926 32213) or donkey-anti mouse (926 32212) secondary antibodies purchased from Licor, all at 1:1,000 dilution in TBS Tween 0.1%. The infrared scanning was performed using the Licor Odyssey scanner (channel, 800; brightness, 50; contrast, 50; sensitivity, auto; resolution, 169.492 µm; pixel area, 0.02873; intensity, 5) and acquired using Odyssey software version 3. Images were then exported as TIFF and cropped using Adobe Photoshop CS4.

Cell proliferation. shPKM and control HCT116, HT29 or SW620 cells were seeded into a 96-well plate at a density of 1,000 cells per well in 200 µl of

DMEM containing 2% FBS. On days 4 to 8 after seeding, 20 µl of Alamar Blue solution (life technologies) was added to each well measured. Cells were incubated for 6 h and fluorescence was measured using 535 nm excitation and 590-nm emission wavelengths.

ATP levels. 6 × 10⁵ cells were seeded on a 6-well plate the day before the experiment. Cells were then washed twice with PBS in order to remove dead cells, and then lysed using the ATP-release buffer (Sigma). ATP was then measured using a luciferase-based assay according to the manufacturer's instructions using the ATP bioluminescent somatic cell assay kit FLASC (Sigma). Values were normalized to the total protein content of the cell lysate as measured by BCA assay (Thermoscientific) using BSA as standard.

Measurement of oxygen-consumption rate and extracellular-acidification rate. 3 × 10⁴ cells were plated onto XF24 plates in DMEM (10% FBS, 2 mM glutamine) (Seahorse Bioscience) and incubated at 37 °C, 5% CO₂ overnight. The medium was then replaced with 675 µl of unbuffered assay media (Seahorse Bioscience) supplemented with 2 mM glutamine, 25 mM glucose and 2% FBS (pH was adjusted to 7.4 using sodium hydroxide 0.5 mM) and cells were then placed at 37 °C in a CO₂-free incubator for 30 min. Basal oxygen-consumption rate (OCR) and extracellular-acidification rate (ECAR) were recorded using the XF24 plate reader. At the end of the experiment 1 µM antimycin A was added in order to measure mitochondria-independent oxygen consumption. Each measurement cycle consisted of 3 min mixing, 3 min waiting and 4 min measuring. OCR and ECAR were normalized to cell number. To obtain the mitochondrial-dependent OCR, only antimycin-A-sensitive respiration was used. Homogeneous plating and cell count were assessed by fixing the cells with 10% trichloroacetic acid for 1 h at 4 °C and then staining the fixed cells with 0.47% solution of Sulforhodamine B (Sigma).

Quantification of intra- and extracellular metabolites by the standard addition method. 1 × 10⁶ cells were plated onto 6-cm plates in triplicates and cultured in standard medium (DMEM, 10% FBS, 2 mM glutamine). Two additional plates were grown as counter plates. The medium was replaced after 24 h by 10 ml of fresh standard medium, and cells were incubated for another 24 h before extraction (as described in the following section). Standard compounds were weighed separately and dissolved together in water to make solution A (in which each metabolite has a concentration between 1 mM and 10 mM). 1 ml of solution A was added to 49 ml of dilution solvent (50:50 acetonitrile:water) to make stock solution B (in which each metabolite had a concentration between 20 µM and 200 µM). For quantification, cells or media extracts (200 µl) were mixed with 800 µl of dilution solvent, containing 0, 4, 20, 100, 300 or 500 µl of stock solution B. Dilutions were analysed by LC-MS. The concentration of each metabolite in the extract was calculated according to the linear regression fit²⁵. All dilution series were performed in triplicates using three biological replicates.

Measurement of ¹³C-labelled metabolites by LC-MS. 4 × 10⁵ cells were plated onto 6-well plates and cultured in standard medium for 24 h. The medium was then replaced by 2 ml of fresh medium containing 5 mM unlabelled glucose and 3 h later 5 mM of U-¹³C-glucose (Cambridge Isotope Laboratories) was added; alternatively, medium was replaced by 2 ml of fresh medium with U-¹³C-glucose only. Cells were incubated for the indicated time before extraction. For extraction, cells were washed twice in PBS and metabolites were extracted on a dry ice/methanol bath in a 50:30:20 ratio of methanol:acetonitrile:water and quickly scraped. The insoluble material was spun down in a cooled centrifuge at 16,000g for 15 min at 0 °C and the supernatant was collected for subsequent LC-MS analysis. The volume of extraction solution was calculated according to cell number and, extrapolated using a 'counter dish' cultured under the same experimental conditions as the sample dishes. A volume of 1 ml of extraction solutions per 2 × 10⁶ cells was used. Metabolites were separated using a liquid chromatography system. A ZIC-pHILIC column (4.6 mm × 150 mm, guard column 4.6 mm × 10 mm; Merck) was used for liquid chromatography separation using gradient elution with a solution of 20 mM ammonium carbonate, with 0.1% ammonium hydroxide, and acetonitrile. Detection of metabolites was performed using a Thermo Scientific Exactive high-resolution mass spectrometer with electrospray ionization, examining metabolites in both positive and negative ion modes, over the mass range of 75–1,000 m/z.

2-dimensional gel electrophoresis. shPKM and control HCT116 cells were lysed precipitated and resuspended as described in ref. 9. Isoelectric focusing (IEF) was performed using ZOOM strips (pH 3–10, non-linear) according to the manufacturer's instructions (Invitrogen). After IEF, strips were equilibrated in buffer containing 10 mg ml⁻¹ DTT and subsequently in buffer containing 25 mg ml⁻¹ iodoacetamide before SDS-PAGE. As negative control, shPKM HCT116 cells were lysed and precipitated in acetone and the pH adjusted to pH 5 by the addition of 1% acetic acid. This removed any acid-labile phosphates such as phosphohistidine. PGAM1 and its phosphorylated forms were detected by western blot

using goat anti-PGAM1 (Novus) 1:1,000 and Donkey anti-Goat (Licor) 1:1,000 antibodies.

PKM2 activator. Cells were treated for 1 h with 20 μ M of the commercially available PKM2 activator 1-(3-chloro-5-trifluoromethyl-pyridin-2-yl)-1H-pyrrole-2-sulfonic acid p-tolylamide²⁶ (indicated in this letter as Cmpd1) or with vehicle as control. Both were incubated in media containing 25 mM U-¹³C-glucose for the indicated time.

In vitro measurement of pyruvate kinase activity. PKM2 was expressed and purified as described in the 'PKM2 X-ray crystallography' section below. Rabbit muscle PKM1 was obtained from Sigma; PKLR was purchased from Abcam. Enzyme activity was measured *in vitro* with a coupled assay quantifying levels of ATP using luminescent Kinase-Glo Plus reagent (Promega) as described previously in ref. 27 with some modifications. Measurements were performed in the presence of 50 mM Tris, pH 7.5, 100 mM KCl, 10 mM MgCl₂, 200 μ M PEP, 200 μ M ADP, 3% DMSO and either 4 nM PKM2, 4 nM PKM1 or 10 nM PKLR. Reactions (25 μ l) were incubated for 20 min on a shaker before addition of 25 μ l Kinase-Glo Plus reagent, as per the manufacturer's instructions. Luminescence was read with a PHERAstar (BMG LABTECH). The signal was normalized to the no-enzyme controls. Activation curves were fitted to a four-parameter logistic equation and K_m curves were fitted to a Michaelis–Menten equation using Prism 5 (GraphPad).

Isothermal titration calorimetry. Isothermal titration calorimetry experiments were performed on a MicroCal VP-ITC at 25 °C in a buffer comprising 50 mM Tris, 100 mM KCl, 10 mM MgCl₂ and 1 mM Tris(carboxyethyl phosphine) (TCEP) at pH 7.5. For titrations the L-serine concentration was 5 mM in the injection syringe and the PKM2 concentration was 28 μ M in the sample cell. The protein concentration refers to the monomer. PKM2 was incubated for 30 min with an excess of FBP (200 μ M) before the L-serine titration was performed in the presence of FBP. The K_d value for L-serine binding was much greater than the PKM2 concentration used, making it difficult to accurately determine the stoichiometry value. Therefore, the stoichiometry parameter was fixed at 1 for the purpose of data analysis using the single site-binding model in Origin 7.0.

PKM2 X-ray crystallography. A publicly available human PKM2 expression construct was obtained from the Structural Genomics Consortium (SGC). His₆-PKM2 was purified using nickel-nitrilotriacetic acid affinity capture and HiLoad Superdex 16/60 S75 size-exclusion chromatography. Human PKM2 was crystallized using hanging-drop vapour diffusion. Protein solution (10 mg ml⁻¹, 25 mM Tris/HCl, pH 7.5, 0.1 M KCl, 5 mM MgCl₂, 10% (v/v) glycerol) was mixed in a 1:1 ratio with reservoir solution containing 0.1 M KCl, 0.2 M ammonium tartrate, 24% (w/v) PEG3350. Crystals were soaked overnight in a solution containing 30 mM L-serine, cryo-protected and flash-frozen in liquid nitrogen. X-ray diffraction data were collected from a single crystal at 100K at Beamline-I03 at the Diamond Light Source. Diffraction data were processed using XDS AutoPROC from Global Phasing and SCALA (CCP4)²⁸. Molecular replacement was performed using model 3H6O (SGC) in CSEARCH²⁹ and maximum-likelihood refinement carried out using a mixture of automated (see ref. 18 and references therein) and manual refinement protocols employing Refmac (CCP4) and AutoBuster from Global Phasing. Ligand fitting was performed using Autosolve²⁹ and manual rebuilding. Simulated annealing was not used. The four PKM2 monomers comprising the tetramer in the asymmetric unit were refined as independent entities, but non-crystallographic symmetry (NCS) restraints were imposed in AutoBuster using the 'ncsauto' command. Refinement of the structure in the absence of NCS restraints gave (R_f = 23.7, R = 17.7) and with 'ncsauto' gave (R_f = 22.7, R = 17.9), showing a small, but significant, reduction in R_f using 'ncsauto' restraints. At the 'effective resolution' of 2.36 Å there are ~86,000 unique reflections. The refinement included ~16,600 non-hydrogen atoms. B -factors were refined isotropically giving a total of ~66,500 parameters for all non-hydrogen atoms in the PKM2 tetramer. The four serine molecules were refined as independent ligands. Details of ligand occupancies, B -factors and so on are tabulated in the crystallographic data tables (Supplementary Tables 1 and 2)^{30–33}.

PKM2 mutagenesis. Human PKM2 point mutant constructs were generated using a Stratagene QuikChange II site directed mutagenesis kit (no. 200524). PCR protocols were as defined in the product manual. The following forward DNA primers, and their reverse complemented primer counterparts, were used for the mutagenesis reactions (sequence of mutated bases shown in uppercase bold): H464A, 5' gctcgtcaggccGCcctgtaccgtggc3'; S437Y, 5' accaagtctggcaggtAtgct caccaggtgg3'. Primers were purchased from Sigma. The previously described SGC human PKM2 construct was used as the DNA template within the PCR reactions.

The presence of the point mutations was confirmed by DNA sequencing of the DNA constructs (Beckman Coulter Genomics Inc.) and in-house LC–MS of the purified recombinant proteins. The mutant proteins were expressed and purified identically to the wild-type protein.

Statistical analyses. The data (mean \pm s.e.m.) are representative of three independent experiments, performed in triplicates (unless otherwise indicated). Data were analysed and presented with GraphPad Prism 5.01 software (GraphPad Software Inc.). P values were calculated using an unpaired two-tailed t -test.

Data processing. The data-processing workflow started by first converting the vendor specific raw data files from the mass spectrometer into the mzXML open-data format³⁴, using the msconvert utility from the ProteoWizard Library and Tools collection³⁵ (<http://proteowizard.sourceforge.net/>). The set of all chromatographic peaks in each of the converted raw files were then extracted using the CentWave³⁶ feature detection algorithm from XCMS³⁷. The resulting data were stored in the PeakML file format³⁸, and the rest of the processing was handled by the scriptable mass spectrometry data-processing tool mzmatch.R³⁹ (<http://mzmatch.sourceforge.net/>).

The next step in the workflow involved aligning and combining the chromatographic features between biological replicates of a single sample. The PeakML files thus created were subjected to an additional filtering procedure to discard all peaks that were not reproducibly detected in all biological replicates involved. Chromatographic peaks of individual samples were then aligned together on the basis of their retention time and m/z values, and combined into a single PeakML file. Peak sets that do not include peaks from every sample were filled in by extracting ion chromatograms within the retention time and mass window of the corresponding peak set directly from the raw data files. From these peak sets, only those that had more peaks than the number of replicates minus one were selected for further analysis. Putative identification of the peak sets were made by matching the detected masses to that of the compounds relevant to this study. Isotope peaks were extracted by identifying the peaks that fell in the retention time window of the identified unlabelled peak and correspond to the estimated mass window (2 p.p.m.) of the isotope. All isotope identification and quantification of the ratios were performed by the PeakML.Isotope.TargettedIsotopes() function of the mzmatch.R library, and detailed documentation and tutorials for which are available at <http://www.mzmatch.sourceforge.net/isotopes-targetted.php>.

25. Luo, B., Groenke, K., Takors, R., Wandrey, C. & Oldiges, M. Simultaneous determination of multiple intracellular metabolites in glycolysis, pentose phosphate pathway and tricarboxylic acid cycle by liquid chromatography-mass spectrometry. *J. Chromatogr. A* **1147**, 153–164 (2007).
26. Salituro, F. G. & Saunders, J. O. Therapeutic compositions and related methods of use. PTC patent application WO/2010/118063 (2010).
27. Boxer, M. B. *et al.* Evaluation of substituted *N,N'*-diarylsulfonamides as activators of the tumor cell specific M2 isoform of pyruvate kinase. *J. Med. Chem.* **53**, 1048–1055 (2010).
28. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
29. Mooij, W. T. *et al.* Automated protein-ligand crystallography for structure-based drug design. *Chem Med Chem* **1**, 827–838 (2006).
30. Brünger, A. T. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472–475 (1992).
31. Diederichs, K. & Karplus, P. A. Improved R factors for diffraction data analysis in macromolecular crystallography. *Nature Struct. Biol.* **4**, 269–275 (1997).
32. Weiss, M. S. & Hilgenfeld, R. On the use of the merging R factor as a quality indicator for X-ray data. *J. Appl. Crystallogr.* **30**, 203–205 (1997).
33. Weiss, M. S. Global indicators of X-ray data quality. *J. Appl. Crystallogr.* **34**, 130–135 (2001).
34. Pedrioli, P. G. *et al.* A common open representation of mass spectrometry data and its application to proteomics research. *Nature Biotechnol.* **22**, 1459–1466 (2004).
35. Kessner, D., Chambers, M., Burke, R., Agus, D. & Mallick, P. ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics* **24**, 2534–2536 (2008).
36. Tautenhahn, R., Bottcher, C. & Neumann, S. Highly sensitive feature detection for high resolution LC/MS. *BMC Bioinformatics* **9**, 504 (2008).
37. Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R. & Siuzdak, G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal. Chem.* **78**, 779–787 (2006).
38. Fernandez, C. A., Des Rosiers, C., Previs, S. F., David, F. & Brunengraber, H. Correction of ¹³C mass isotopomer distributions for natural stable isotope abundance. *J. Mass Spectrom.* **31**, 255–262 (1996).
39. Scheltema, R. A., Jankevics, A., Jansen, R. C., Swertz, M. A. & Breitling, R. PeakML/mzMatch: a file format, Java library, R library, and tool-chain for mass spectrometry data analysis. *Anal. Chem.* **83**, 2786–2793 (2011).

A vaccine strategy that protects against genital herpes by establishing local memory T cells

Haina Shin¹ & Akiko Iwasaki¹

Most successful existing vaccines rely on neutralizing antibodies, which may not require specific anatomical localization of B cells. However, efficacious vaccines that rely on T cells for protection have been difficult to develop, as robust systemic memory T-cell responses do not necessarily correlate with host protection¹. In peripheral sites, tissue-resident memory T cells provide superior protection compared to circulating memory T cells^{2,3}. Here we describe a simple and non-inflammatory vaccine strategy that enables the establishment of a protective memory T-cell pool within peripheral tissue. The female genital tract, which is a portal of entry for sexually transmitted infections, is an immunologically restrictive tissue that prevents entry of activated T cells in the absence of inflammation or infection⁴. To overcome this obstacle, we developed a vaccine strategy that we term 'prime and pull' to establish local tissue-resident memory T cells at a site of potential viral exposure. This approach relies on two steps: conventional parenteral vaccination to elicit systemic T-cell responses (prime), followed by recruitment of activated T cells by means of topical chemokine application to the restrictive genital tract (pull), where such T cells establish a long-term niche and mediate protective immunity. In mice, prime and pull protocol reduces the spread of infectious herpes simplex virus 2 into the sensory neurons and prevents development of clinical disease. These results reveal a promising vaccination strategy against herpes simplex virus 2, and potentially against other sexually transmitted infections such as human immunodeficiency virus.

Viral sexually transmitted infections (STIs) such as human immunodeficiency virus 1 (HIV-1) and herpes simplex virus 2 (HSV-2) account for considerable morbidity and mortality around the world. Strong preclinical evidence for the role of T cells in controlling viral STIs has led to the design of prophylactic vaccines that elicit systemic cellular immunity, and yet these vaccines have not been efficacious^{1,5}. Although systemic memory T cells can migrate freely through organs such as the spleen and liver, other parts of the body such as the intestines, lung airways, central nervous system, skin and vagina are restrictive for memory T-cell entry⁶. In the latter tissues, inflammation or infection is often required to permit entry of circulating activated T cells to establish a tissue-resident memory T-cell pool that is separate from the circulating pool^{2,7,8}. Given that side effects of inflammation in the reproductive tissue may preclude the use of a live prophylactic vaccine given vaginally, we investigated an alternative approach to recruit virus-specific T cells into the vaginal mucosa without inducing local inflammation or infection.

After genital HSV-2 infection, chemokine (C-X-C motif) ligand 9 (CXCL9) and CXCL10 expression is induced by interferon- γ secreted by CD4⁺ T cells and mediates recruitment of effector CD8⁺ T cells to the infected tissue via the chemokine receptor CXCR3 (ref. 4). CXCR3 is expressed by both effector T-helper 1 (T_H1) cells and activated CD8⁺ T cells, as well as other cell types⁹. Thus, we proposed that the topical application of chemokines CXCL9 and CXCL10 might be sufficient to recruit effector T cells to the vagina in the absence of infection. To test this hypothesis we used T-cell antigen receptor

transgenic CD8⁺ T cells that recognize an epitope within the HSV glycoprotein B (gBT-I)¹⁰ to track the HSV-2 specific CD8 T-cell population. Naive female C57BL/6 mice were transplanted with 10⁵ congenically marked gBT-I CD8⁺ T cells and immunized subcutaneously with an attenuated strain of HSV-2 that lacks thymidine kinase (TK⁻ HSV-2)¹¹ (Fig. 1a). As expected, this route of immunization resulted in minimal migration of activated CD8⁺ T cells into the vagina (Fig. 1b, c). To recruit or 'pull' activated HSV-specific CD8⁺ T cells, the chemokines CXCL9 and CXCL10 were topically applied to the vaginal cavity of subcutaneously immunized mice (Fig. 1a). Another group of mice was immunized intravaginally with TK⁻ HSV-2, which served as a positive control for maximal CD8⁺ T-cell recruitment to the vagina (Fig. 1b, c). At day 6 post infection, all three treatment groups exhibited primary CD8⁺ T-cell responses of similar magnitudes, as indicated by the numbers and percentages of systemic gBT-I CD8⁺ T cells found in the spleen (Fig. 1b, c). However, the number and percentage of gBT-I CD8⁺ T cells in the vagina were significantly higher in mice treated with the chemokine pull (subcutaneous immunization plus pull) compared to the control subcutaneously immunized mice (Fig. 1b, c). Furthermore, the action of the chemokine pull was restricted to the genital mucosa, as gBT-I CD8⁺ T-cell recruitment to the vagina-draining iliac lymph nodes was limited (Fig. 1c). Activated CD4⁺ T cells were also strongly recruited to the vagina by the chemokine pull (Fig. 1d). Antigen in the vagina was not responsible for the recruitment, as HSV-2 genomic DNA was absent from the genital tract after subcutaneous immunization (Supplementary Fig. 1). To mimic a vaccination scenario more closely, we also tested whether endogenous virus-specific T cells could be recruited by prime and pull. Like gBT-I CD8⁺ T cells, the systemic endogenous HSV-specific CD8⁺ T-cell response was similar in all immunized groups (Supplementary Fig. 2a). However, significantly greater numbers of HSV-specific CD8⁺ T cells as well as CD4⁺ T cells were present in the genital tracts of mice treated with the chemokine pull as compared to subcutaneous immunization alone (Supplementary Fig. 2). Thus, these data show that the prime and pull method is capable of recruiting a large number of parenterally primed T cells to the genital tract with a single topical application of chemokines.

To assess the possible inflammatory consequences of topical chemokine application to the vagina, we examined the presence of innate inflammatory cells after the chemokine pull. Other cell types, including natural killer cells and plasmacytoid dendritic cells (pDCs) express CXCR3 (ref. 9). However, no significant increase in the number of pDCs, natural killer cells, granulocytes, dendritic cells, monocytes, macrophages and monocyte-derived dendritic cells was elicited by the chemokine treatment (subcutaneously immunized plus pull) compared to the subcutaneously immunized control (Supplementary Fig. 3). These data indicated that topical chemokines do not induce appreciable recruitment of natural killer cells or pDCs to the vagina and that effector T cells are selectively recruited during prime and pull without inducing a general inflammatory response.

During genital HSV infection, CD4⁺ T cells act as a pioneering population for the migration of virus-specific CD8⁺ T cells by inducing the

¹Department of Immunobiology, Yale University School of Medicine, 300 Cedar Street, New Haven, Connecticut 06520, USA.

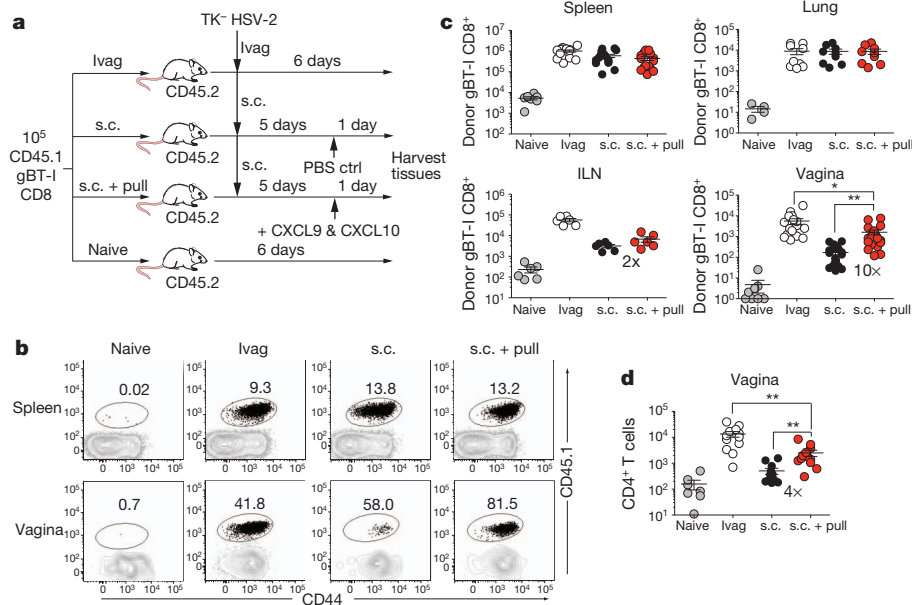


Figure 1 | Effector T cells are recruited to the vagina by topical chemokine treatment. **a**, Experimental schematic. Donor gBT-I CD8⁺ T-cell recipients were not immunized (naive), or were immunized either intravaginally (ivag) or subcutaneously (s.c.) with TK⁻ HSV-2. Five days post infection, subcutaneously immunized mice were treated vaginally with either the chemokines CXCL9 and CXCL10 (pull) or PBS. **b**, The frequency of donor gBT-I CD8⁺ T cells 1 day post pull in the indicated tissues (ILN, iliac lymph nodes). Plots are gated on total CD8⁺ T cells and numbers indicate per cent gBT-I (CD45.1⁺). **c**, The number of donor gBT-I CD8⁺ T cells 1 day post pull

in the indicated tissues. **d**, The number of CD4⁺ T cells 1 day post pull in the vagina. **c**, **d**, Numbers in graphs indicate fold difference in T-cell number for subcutaneously immunized compared with subcutaneously immunized plus pull. Statistical significance was determined by two-tailed unpaired Student's *t*-test. Data are pooled from 2–7 independent experiments (*n* = 6–21 per group). All error bars show s.e.m. and **P* = 0.05–0.01, ***P* = 0.01–0.001, ****P* < 0.001 throughout all figures (individual *P* values are found in the Supplementary Information).

production of critical chemokines within the tissue⁴. To determine whether the recruitment of gBT-I CD8⁺ T cells to the genital tract was similarly dependent on CD4⁺ T-cell help during prime and pull, subcutaneously immunized mice were injected with a CD4-depleting antibody on day 3 and day 5 post infection to preserve normal CD8⁺ T-cell priming¹², and then treated with the chemokine pull (Supplementary Fig. 4a). In CD4⁺ T-cell-depleted mice (Supplementary Fig. 4b), both systemic gBT-I CD8⁺ T-cell numbers and migration to the vagina were unaffected (Supplementary Fig. 4c, d), indicating that recruitment of effector CD8⁺ T cells to the vagina after chemokine treatment bypasses the requirement for CD4⁺ T-cell help.

CXCR3 is upregulated on T cells upon activation and remains high through the effector and memory stages⁹. Having demonstrated that CXCL9 and CXCL10 could recruit CXCR3⁺ effector T cells to the vagina, we next examined the efficacy of the chemokine pull at different stages of T-cell priming. After subcutaneous TK⁻ HSV-2 immunization, CXCR3 was upregulated on both gBT-I CD8⁺ T cells and CD4⁺ T cells throughout the response (Fig. 2a), suggesting that both effector and memory T cells should be capable of responding to the chemokine pull. Previous reports have shown that early effector CD8⁺ T cells had an increased ability to migrate to peripheral tissues⁷, so we next determined whether the timing of chemokine pull dictated the efficacy of T-cell recruitment to the genital tract. When subcutaneously immunized mice were treated with the chemokine pull at the effector (day 5), contraction (day 15) and memory (day 28) phase of the T-cell response¹³, we found that the chemokine pull was most effective at recruiting antigen-specific CD8⁺ T cells during the effector (day 5) phase, which correlated with the increased number and percentage of systemic gBT-I CD8⁺ T cells (Fig. 2b). Despite similar CXCR3 expression (Fig. 2a), memory gBT-I CD8⁺ T cells were not present in the tissue after pull when treated during the memory phase (day 28) (Fig. 2b). We speculate that this might be due to altered homing patterns^{7,14} and the reduced number and percentage of gBT-I CD8⁺ T cells in circulation at the memory time point. Recruitment of CD44⁺ CD4⁺ T cells

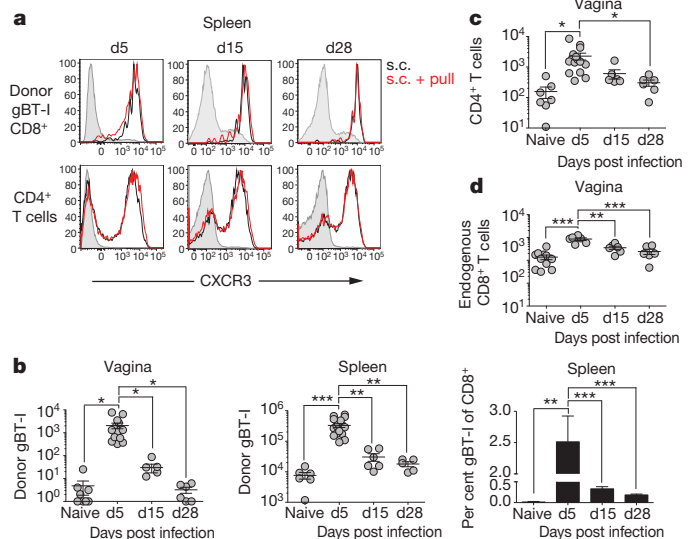


Figure 2 | Chemokine pull is specific for highly activated effector T cells. Mice were subcutaneously immunized and given chemokines or PBS at day 5, 15 or 28 post infection and analysed 1 day post pull. **a**, CXCR3 expression on donor gBT-I CD8⁺ T cells or CD44⁺ CD4⁺ T cells from the spleen 1 day post pull from subcutaneously immunized mice (open black) and subcutaneously immunized plus pull (open red). Shaded histograms are CD44^{lo} CD8⁺ or total CD4⁺ T cells. **b**, The gBT-I CD8⁺ T cell number in the vagina (left) or spleen (middle) and frequency in the spleen (right) were examined 1 day post pull. **c**, The number of CD44⁺ CD4⁺ T cells in the vagina 1 day post pull on the indicated days post infection. **d**, The number of endogenous CD44⁺ CD8⁺ T cells in the vagina 1 day post pull on the indicated days post infection. **b**–**d**, Statistical significance was determined by two-tailed unpaired Student's *t*-test. Data are pooled from 2–3 independent experiments; *n* = 6–18 per group.

(Fig. 2c) and endogenous CD8⁺ T cells (Fig. 2d) followed a similar pattern. Collectively, these data indicate that the chemokine pull is most effective at recruiting recently activated effector CD8⁺ T cells that are circulating at high frequency, establishing a specific time frame within which the chemokine pull should be administered after priming.

For the prime and pull approach to be an effective vaccine strategy, pathogen-specific T cells must be retained within the tissue for an extended time and establish a pool of memory cells. To determine whether the effector gBT-I CD8⁺ T cells recruited into the vagina after prime and pull were capable of establishing a long-term population of memory CD8 T cells, we examined the presence of gBT-I CD8⁺ T cells 4 weeks after the chemokine pull. The number of systemic memory gBT-I CD8⁺ T cells, although decreased compared to day 1 post pull (Fig. 1c) due to contraction of the T-cell response, was similar regardless of immunization route or treatment (Fig. 3a). However, a significantly greater number and percentage of memory gBT-I CD8⁺ T cells was present in the genital tract of subcutaneously immunized mice treated with the chemokine (subcutaneously immunized plus pull) compared to chemokine untreated mice (subcutaneously immunized) (Fig. 3a, b). Despite significant recruitment during the effector phase (Fig. 1d), CD4⁺ T cells were not retained within the vagina long term (Fig. 3c), reminiscent of CD4⁺ T-cell behaviour after dermal HSV-1 infection in which the cells leave the site of infection to mediate immunosurveillance¹⁵. Thus, CD4⁺ T cells may require additional signals, such as those generated during HSV-2 infection^{16,17}, to be retained long term within the vagina.

To investigate the stability of this tissue-resident population of memory gBT-I CD8⁺ T cells, we also examined T-cell numbers at 12 weeks post pull. Donor gBT-I CD8⁺ T-cell numbers in the vagina were significantly higher after prime and pull than after subcutaneous immunization alone (Fig. 3d). Furthermore, the number of memory gBT-I CD8⁺ T cells did not decline between 4 weeks and 12 weeks (Fig. 3d), suggesting that this tissue-resident population was stable and retained long term. CD4⁺ T-cell numbers in the vagina remained low at week 12 after prime and pull, and were comparable to numbers

detected at week 4 (Fig. 3e). Thus, a single chemokine pull given to mice during the effector phase is sufficient to establish a long-term population of tissue-resident memory CD8⁺ T cells, but not CD4⁺ T cells, within the vagina.

Tissue-resident memory T cells are effective in mediating immunity against local infections^{2,3}. HSV-2 spreads from its initial replication site at the epithelium to the innervating neurons, and subsequently establishes latency within the dorsal root ganglia (DRG)⁵. Reactivation from latency leads to viral shedding and formation of genital lesions that are commonly associated with genital herpes⁵. Thus, preventing the spread of virus from the mucosal epithelium to the DRG is key in preventing disease and transmission of the virus. As a single chemokine pull administered after subcutaneous immunization is capable of establishing a population of tissue-resident memory CD8⁺ T cells within the vagina long term, we next examined whether the prime and pull strategy would provide enhanced immunity against genital HSV-2 infection. Mice were challenged intravaginally with a lethal dose of wild-type HSV-2 4 weeks after prime and pull and monitored for disease and survival. Notably, mice treated with the chemokine pull (subcutaneously immunized plus pull) lost significantly less weight than either the non-immunized or subcutaneously immunized controls (Fig. 4a). Furthermore, prime and pull almost completely prevented the development of clinical symptoms, which were observed in both non-immunized and subcutaneously immunized controls (Fig. 4b). Accordingly, mice treated with the chemokine pull had a 100% survival rate compared to the 36.3% survival rate of the subcutaneously immunized control (Fig. 4c). Upon challenge with wild-type HSV-2 4 weeks post pull, mice immunized and chemokine-treated in the absence of T-cell antigen receptor transgenic CD8⁺ T cells were also significantly protected from weight loss (Supplementary Fig. 5a) and clinical disease (Supplementary Fig. 5b), although we did not observe a significant difference in survival rate (Supplementary Fig. 5c). Anti-HSV antibody titres in the vagina were not significantly different between subcutaneously immunized controls and chemokine-treated mice (Supplementary Fig. 6), suggesting that the control of viral challenge was

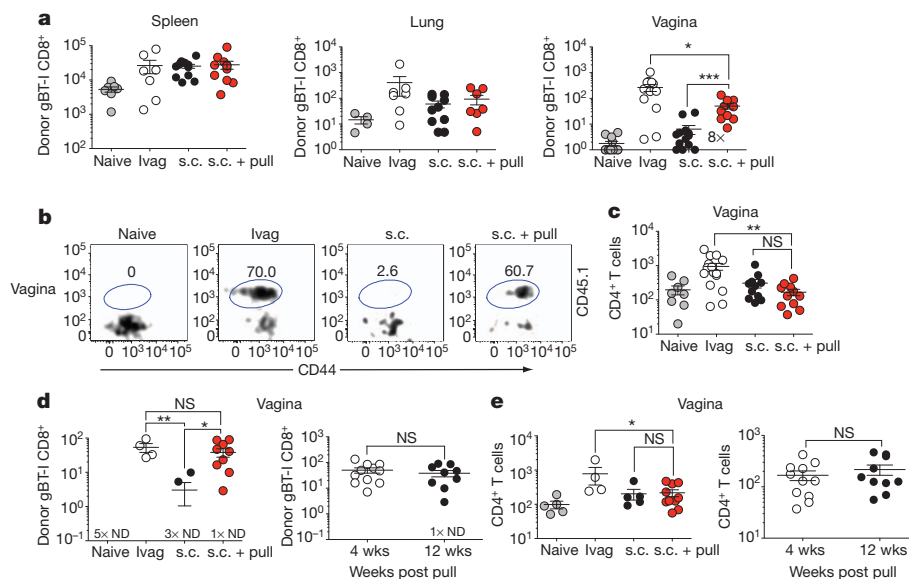


Figure 3 | Virus-specific T cells recruited by chemokine pull are retained in the vagina long term. **a**, Mice were immunized and treated as shown in Fig. 1a. At 4 weeks post pull, numbers of gBT-I CD8⁺ T cells were determined in the indicated tissues. The number inside the graph shows fold difference in gBT-I number between subcutaneously immunized compared to subcutaneously immunized plus pull groups. **b**, Four weeks post pull, the frequency of gBT-I cells was measured in the vagina. Plots are gated on total CD8⁺ T cells. Numbers show per cent gBT-I (CD45.1⁺). **c**, The number of endogenous CD4⁺

T cells in the vagina at 4 weeks post pull. **d**, The numbers of gBT-I cells were determined in the vagina at 12 weeks post pull (left) and compared to numbers at 4 weeks post pull (right). Number \times ND, number of animals with no cells detected in the tissue. **e**, The numbers of CD4⁺ T cells were determined in the vagina at 12 weeks post pull (left) and compared to the corresponding numbers at 4 weeks post pull (right). Statistical significance was determined by two-tailed unpaired Student's *t*-test. NS, not significant. Data are pooled from 2–3 independent experiments; *n* = 4–15 per group).

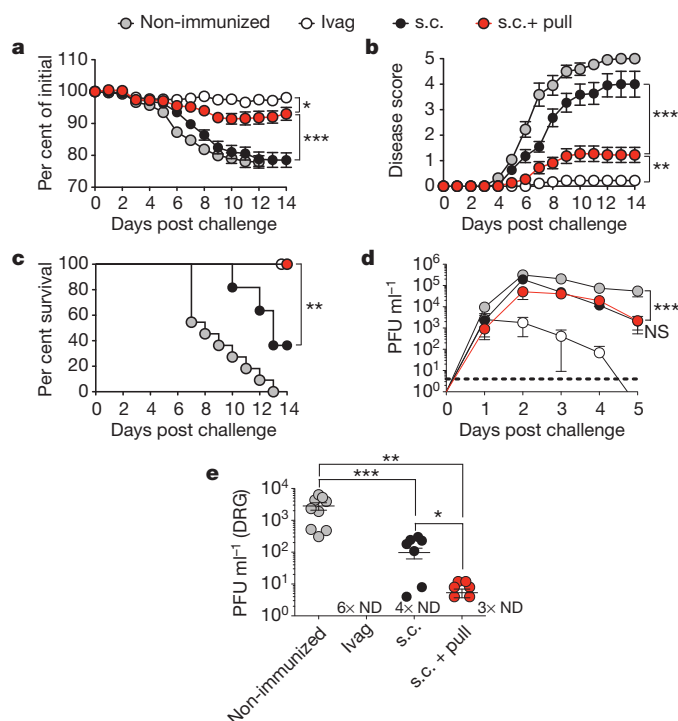


Figure 4 | Prime and pull protects mice from lethal genital HSV-2 challenge.

a, Weight loss in mice immunized as shown in Fig. 1a and then challenged vaginally with a lethal dose of HSV-2 4 weeks post pull. **b**, Disease severity in mice immunized as shown in Fig. 1a and then challenged vaginally with a lethal dose of HSV-2 four weeks post pull. A higher disease score indicates more severe disease symptoms. **c**, Survival in mice immunized as shown in Fig. 1a and then challenged vaginally with a lethal dose of HSV-2 4 weeks post pull. **d**, HSV-2 viral titres from vaginal washes collected at the indicated time points post challenge with HSV-2. Dashed line indicates limit of detection, none detected. $n = 11$ (non-immunized), $n = 9$ (intravaginal immunization), $n = 12$ (subcutaneously immunized control, subcutaneously immunized plus pull). **e**, Viral titres were measured in the dorsal root ganglia 6–7 days post challenge. Number \times ND, number of mice in which no virus was detected. $n = 6$ –11 per group. Statistical significance was measured by two-way ANOVA (**a**, **b**, **d**), log-rank (Mantel-Cox) test (**c**) or two-tailed unpaired Student's *t*-test (**e**). Data are pooled from 3–5 independent experiments.

probably T-cell mediated. These results demonstrate that the addition of a chemokine pull to parenteral immunization could greatly enhance protective immunity against genital HSV-2 infection.

To test more stringently the long-term protection afforded by prime and pull, we also challenged mice 10–12 weeks post pull. At this late time point, the prime and pull group lost less weight compared to subcutaneously immunized controls (Supplementary Fig. 7a), and were significantly protected from development of disease (Supplementary Fig. 7b). Furthermore, at 2 weeks post challenge, the prime and pull group had a survival rate of 100%, whereas subcutaneously immunized controls had a survival rate of 57% (Supplementary Fig. 7c). Thus, our results show that the protection provided by prime and pull lasts over time and remains robust up to 12 weeks after chemokine treatment.

To determine the mechanism by which the prime and pull strategy mediates protection from HSV-2 disease, we measured viral replication within the genital mucosa. Notably, we found no difference in virus titres from the vaginal secretion of subcutaneously immunized versus chemokine-treated (subcutaneously immunized plus pull) mice (Fig. 4d and Supplementary Figs. 5d and 7d), indicating that protection was probably being mediated at a different location. As the more severe symptoms of clinical disease in mice are associated with viral replication in the peripheral nervous system¹⁸, we next investigated whether prime and pull could protect the DRG against infection. When viral

replication within the DRG was measured, we found that mice treated with the chemokine pull had significantly lower virus titres than non-immunized mice (Fig. 4e). Furthermore, viral titres in the DRG of the prime and pull group were significantly lower than that of subcutaneously immunized mice (Fig. 4e). Together, these data indicate that prime and pull strategy greatly reduces disease by controlling neuronal infection with HSV-2 rather than by controlling mucosal viral replication.

Our study demonstrates that after conventional vaccination to generate a systemic T-cell population (prime), a single topical treatment with chemokines applied vaginally (pull) can provide superior protection against genital herpes by decreasing the spread of virus from the mucosal epithelia into the neurons. Importantly, protection of neurons from HSV-2 infection by prime and pull may decrease reactivation and viral shedding, which may reduce disease and transmission. Although the exact role of T cells in controlling neuronal HSV-2 infection after prime and pull is not yet clear, we speculate that the local HSV-specific T cells may help to control entry of virus at the neuronal endings, or promote blockade of viral replication once inside the neurons. Furthermore, other studies have demonstrated that T cells recruited to the genital tract by inflammation alone can decrease viral replication at the mucosal surface¹⁹, suggesting that control of infection at the site of entry may be possible by optimizing prime and pull. Thus, in addition to preventing reactivation of latent HSV²⁰, virus-specific memory T cells may be mobilized to control neuronal viral infection during primary infection. Although topical application in the genital tract of Toll-like receptor ligands such as imiquimod have been shown to be effective as a therapeutic approach²¹, they may not be ideal vaccine candidates as they seem to be effective for only a short time after application and function through the induction of pro-inflammatory cytokines²². The novel prime and pull vaccine strategy described here provides an alternative to direct immunization of the genital tract, and establishes robust, long-term immunity with minimal local inflammation.

Cellular immunity is critical in mediating protection against viral STIs such as HSV-2 and HIV-1 (ref. 23). Both viruses enter through the genital mucosa, begin local replication and then spread to other tissues²³. Although our data highlight the role of prime and pull in controlling viral spread to the peripheral nervous system, this vaccination approach is not necessarily restricted to neurotropic viruses. HIV-1 enters the genital mucosa and invades the draining lymph node, from which systemic dissemination of the virus occurs²³. In its current form, prime and pull establishes tissue-resident memory CD8⁺ T cells but not CD4⁺ T cells. Given that a single HIV-1 virion can establish infection in humans²³, local memory CD8⁺ T cells may be key to protection against HIV-1 (ref. 23) by reducing replication and dissemination of the founder virus, while the absence of local CD4⁺ T cells could limit the availability of immediate target cells. Beyond viral infections, prime and pull could be applied to improve recruitment of immune cells to other restrictive microenvironments such as solid tumours. Effective immunotherapy can be hindered by either decreased or inappropriate expression of chemokines at the tumour tissue, leading to minimal migration of immune cells²⁴. Delivery of appropriate chemokines to the tumour tissue after immunization could enhance recruitment of tumour-specific T cells and augment the efficacy of immunotherapies. Although the protocol we present pairs the pull with a suboptimal subcutaneous immunization, we propose that the prime and pull strategy could be used in conjunction with more effective immunizations⁵ to enhance protection. Ultimately, the ability to boost recruitment of T cells and establish resident T-cell populations in peripheral tissues restrictive for lymphocyte homing will aid not only in the prevention but also in the treatment of a wide variety of diseases.

METHODS SUMMARY

Adoptive transfers, infections and T-cell depletion. CD8⁺ T cells (10^5) from the spleens of naive CD45.1⁺ gBT-I TCR transgenic mice¹⁰ were adoptively

transferred into Depo-Provera (GE Healthcare) treated²⁵, naive 6-week-old C57BL/6 recipients (National Cancer Institute). Recipients were immunized intravaginally or subcutaneously in the neck ruff with 10^5 or 10^6 plaque forming units (PFU) of 186TKΔkpn HSV-2 (TK⁻ HSV-2)¹¹, respectively. Some mice were treated twice with 200 μg anti-CD4 antibody (GK1.5) intraperitoneally to deplete CD4⁺ T cells. Five days post-immunization, subcutaneously immunized mice were vaginally swabbed with a Calginate swab (Fisher) and either PBS or a solution of CXCL9 and CXCL10 (3 μg each, Peprotech) in PBS was delivered via pipette tip into the vagina. For 4-week challenges, mice were infected intravaginally with 5,000 PFU of wild-type HSV-2 186 syn+ (ref. 26). For 10–12 week challenges, mice were treated again with Depo-Provera 9–10 weeks before challenge.

Full Methods and any associated references are available in the online version of the paper.

Received 16 January; accepted 17 August 2012.

Published online 17 October 2012.

- McElrath, M. J. & Haynes, B. F. Induction of immunity to human immunodeficiency virus type-1 by vaccination. *Immunity* **33**, 542–554 (2010).
- Gebhardt, T. *et al.* Memory T cells in nonlymphoid tissue that provide enhanced local immunity during infection with herpes simplex virus. *Nature Immunol.* **10**, 524–530 (2009).
- Jiang, X. *et al.* Skin infection generates non-migratory memory CD8⁺ T_{RM} cells providing global skin immunity. *Nature* **483**, 227–231 (2012).
- Nakanishi, Y., Lu, B., Gerard, C. & Iwasaki, A. CD8⁺ T lymphocyte mobilization to virus-infected tissue requires CD4⁺ T-cell help. *Nature* **462**, 510–513 (2009).
- Koelle, D. M. & Corey, L. Herpes simplex: insights on pathogenesis and possible vaccines. *Annu. Rev. Med.* **59**, 381–395 (2008).
- Woodland, D. L. & Kohlmeier, J. E. Migration, maintenance and recall of memory T cells in peripheral tissues. *Nature Rev. Immunol.* **9**, 153–161 (2009).
- Masopust, D. *et al.* Dynamic T cell migration program provides resident memory within intestinal epithelium. *J. Exp. Med.* **207**, 553–564 (2010).
- Klonowski, K. D. *et al.* Dynamics of blood-borne CD8 memory T cell migration *in vivo*. *Immunity* **20**, 551–562 (2004).
- Groom, J. R. & Luster, A. D. CXCR3 ligands: redundant, collaborative and antagonistic functions. *Immunol. Cell Biol.* **89**, 207–215 (2011).
- Mueller, S. N., Heath, W., McLain, J. D., Carbone, F. R. & Jones, C. M. Characterization of two TCR transgenic mouse lines specific for herpes simplex virus. *Immunol. Cell Biol.* **80**, 156–163 (2002).
- Jones, C. A., Taylor, T. J. & Knipe, D. M. Biological properties of herpes simplex virus 2 replication-defective mutant strains in a murine nasal infection model. *Virology* **278**, 137–150 (2000).
- Smith, C. M. *et al.* Cognate CD4⁺ T cell licensing of dendritic cells in CD8⁺ T cell immunity. *Nature Immunol.* **5**, 1143–1148 (2004).
- Kaech, S. M. & Wherry, E. J. Heterogeneity and cell-fate decisions in effector and memory CD8⁺ T cell differentiation during viral infection. *Immunity* **27**, 393–405 (2007).
- Weninger, W., Crowley, M. A., Manjunath, N. & von Andrian, U. H. Migratory properties of naive, effector, and memory CD8⁺ T cells. *J. Exp. Med.* **194**, 953–966 (2001).
- Gebhardt, T. *et al.* Different patterns of peripheral migration by memory CD4⁺ and CD8⁺ T cells. *Nature* **477**, 216–219 (2011).
- Zhu, J. *et al.* Persistence of HIV-1 receptor-positive cells after HSV-2 reactivation is a potential mechanism for increased HIV-1 acquisition. *Nature Med.* **15**, 886–892 (2009).
- Iijima, N. *et al.* Dendritic cells and B cells maximize mucosal Th1 memory response to herpes simplex virus. *J. Exp. Med.* **205**, 3041–3052 (2008).
- Parr, M. B. & Parr, E. L. Intravaginal administration of herpes simplex virus type 2 to mice leads to infection of several neural and extraneural sites. *J. Neurovirol.* **9**, 594–602 (2003).
- Mackay, L. K. *et al.* Long-lived epithelial immunity by tissue-resident memory T (TRM) cells in the absence of persisting local antigen presentation. *Proc. Natl Acad. Sci. USA* **109**, 7037–7042 (2012).
- Khanna, K. M., Lepisto, A. J. & Hendricks, R. L. Immunity to latent viral infection: many skirmishes but few fatalities. *Trends Immunol.* **25**, 230–234 (2004).
- Perkins, N., Nisbet, M. & Thomas, M. Topical imiquimod treatment of aciclovir-resistant herpes simplex disease: case series and literature review. *Sex. Transm. Infect.* **87**, 292–295 (2011).
- Gill, N., Davies, E. J. & Ashkar, A. A. The role of Toll-like receptor ligands/agonists in protection against genital HSV-2 infection. *Am. J. Reprod. Immunol.* **59**, 35–43 (2008).
- Iwasaki, A. Antiviral immune responses in the genital tract: clues for vaccines. *Nature Rev. Immunol.* **10**, 699–711 (2010).
- Gajewski, T. F., Fuentes, M., Spaapen, R., Zheng, Y. & Kline, J. Molecular profiling to identify relevant immune resistance mechanisms in the tumor microenvironment. *Curr. Opin. Immunol.* **23**, 286–292 (2011).
- Parr, M. B. *et al.* A mouse model for studies of mucosal immunity to vaginal infection by herpes simplex virus type 2. *Lab. Invest.* **70**, 369–380 (1994).
- Spang, A. E., Godowski, P. J. & Knipe, D. M. Characterization of herpes simplex virus 2 temperature-sensitive mutants whose lesions map in or near the coding sequences for the major DNA-binding protein. *J. Virol.* **45**, 332–342 (1983).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank E. Foxman and R. Medzhitov for critical reading of the manuscript, and N. Iijima, H. Dong and B. Yordy for technical support. H.S. is supported by NIAID grant F32AI091024. This work is supported by NIH grants AI054359 and AI062428 to A.I.

Author Contributions Experiments were conceived and designed by H.S. and A.I. Experiments were performed by H.S. Data were analysed by H.S. and A.I. The paper was written by H.S. and A.I.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.I. (akiko.iwasaki@yale.edu).

METHODS

Mice. Female 6-week-old C57BL/6 mice were purchased from the National Cancer Institute. gBT-I T cell antigen receptor (TCR) transgenic mice specific for the glycoprotein B epitope gB(498–505) were provided by F. R. Carbone and W. R. Heath and bred in our facility to C57BL/6-Ly5.2Cr mice (CD45.1⁺) (National Cancer Institute). All procedures used in this study complied with federal and institutional policies of the Yale Animal Care and Use committee.

Adoptive transfers and infections. Spleens were collected from naive CD45.1⁺ gBT-I TCR transgenic mice and CD8⁺ T cells were magnetically purified by CD8 α microbeads or CD8 α T-cell isolation kits (Miltenyi Biotec). Donor cells (10⁵) gBT-I CD8⁺ T cells were adoptively transferred into Depo-Provera-treated (GE Healthcare), 7–8-week-old C57BL/6 recipients retro-orbitally. Mice were then immunized intravaginally or subcutaneously with 10⁵ or 10⁶ plaque forming units (PFU) of 186TK Δ kpn HSV-2 (TK⁻ HSV-2) respectively. At 5 days post-infection, the vaginal cavity of mice was swabbed with a Calginate swab (Fisher) and either PBS or a solution of CXCL9 and CXCL10 (3 μ g each, Peprotech) in PBS was delivered via pipette tip into the vagina. Where indicated, C57BL/6 mice that did not receive gBT-I cells were primed and pulled in a similar manner. Some subcutaneously immunized mice were intraperitoneally injected with 200 μ g anti-CD4 (GK1.5) antibody at day 3 and 5 post infection to deplete CD4⁺ T cells. For the 4-week challenge, non-immunized or previously immunized mice at the indicated time points were infected intravaginally with 5,000 PFU wild-type HSV-2 186 syn⁺. Challenges given at 10–12 weeks post pull were treated with Depo-Provera for a second time 1–2 weeks post pull (9–10 weeks before challenge) before infection with 5,000 PFU wild-type HSV-2 186 syn⁺.

Flow cytometry. At various time points, single cell suspensions from the spleen, lungs, vagina and iliac lymph nodes were prepared for analysis as described¹⁷. Briefly, lungs were digested with collagenase D (Roche). Vaginas were treated with Dispase II (Roche) for 15 min and then collagenase D for 30 min. Cells from the spleen and iliac lymph node were counted by haemocytometer. Lung and vagina cell numbers were quantified using CountBright absolute counting beads (Invitrogen). Dead cells were excluded from analysis using the LIVE/DEAD Fixable Aqua Dead Cell stain kit (Invitrogen). All samples were acquired on an LSRII equipped with a 532-nm green laser (BD Biosciences). All data were analysed with FlowJo (Treestar).

Antibodies. The following antibodies were used for this study: CD3 (17A2), CD8 (53-6.7), CD44 (1M7), CD45.1 (A20), CXCR3 (CXCR3-173), CD11c (N418), CD11b (M1/70), MHC class II (M5/114.15.2), Ly6G (1A8), F4/80 (BM8), B220 (RA3-6B2), CD19 (ebio1D3) and NK1.1 (PD136) (Biolegend); Ly6C (AL-21) (BD Biosciences); CD4 (RM4-4) (eBioscience); and CD4 (RM4-5) (Biolegend and Invitrogen). H-2K^b-gB_{498–505} tetramer was obtained from the National Institutes of Health tetramer core facility.

Measurement of viral titres, weight and disease scores. Vaginal secretions were collected 5 days post challenge using PBS and Calginate swabs. Lumbar and sacral dorsal root ganglia (DRG) were collected at days 6–7 post challenge as described²⁷. DRG were homogenized using a motorized pestle (VWR). Titres from vaginal and DRG samples were measured on Vero cell monolayers as previously described¹⁷. Weight loss was measured daily and normalized to body weight on day 0 of challenge. Disease was monitored daily and scored as follows: (0) no disease; (1) genital inflammation; (2) genital lesions and hair loss; (3) hunched posture and ruffled fur; (4) hind limb paralysis; and (5) premonitory²⁸. Mice were euthanized before reaching the moribund state due to humane concerns.

Detection of HSV-2 antigen by quantitative PCR. Mice were immunized subcutaneously or intravaginally and were killed at day 5 post infection. Vaginal tissue was collected and genomic DNA was extracted as previously described²⁹. Briefly, tissue was homogenized in a salt homogenizing buffer using a motorized pestle. Proteinase K and SDS were added to samples and incubated overnight at 55 °C. After addition of a sodium chloride solution, samples were centrifuged and supernatants were transferred to new tubes. Isopropanol was added to the supernatants and incubated at 20 °C for 1 h. DNA was pelleted by centrifugation, washed with ethanol and resuspended in H₂O. HSV-2 was measured with primers detecting glycoprotein B (gB) (Forward: 5'-AGACCAGGGCCGCTGATC-3'; reverse: 5'-GCGCTGGACCTCCGTGTAG-3') with quantitative polymerase chain reaction (Stratagene). DNA purified from TK⁻ HSV-2 was used as standard to calculate PFU equivalents.

Measurement of HSV-specific antibody titres. Vaginal secretions were collected from mice with PBS and Calginate swabs 4 weeks post pull. HSV-specific immunoglobulin-G (IgG) was measured by ELISA assay as previously described³⁰. Known quantities of anti-HSV gB monoclonal antibody (SS10 mouse IgG) kindly provided by G. Cohen and R. Eisenberg was used as a standard.

27. Malin, S. A., Davis, B. M. & Molliver, D. C. Production of dissociated sensory neuron cultures and considerations for their use in studying neuronal function and plasticity. *Nature Protocols* **2**, 152–160 (2007).
28. Morrison, L. A., Da Costa, X. J. & Knipe, D. M. Influence of mucosal and parenteral immunization with a replication-defective mutant of HSV-2 on immune responses and protection from genital challenge. *Virology* **243**, 178–187 (1998).
29. Aljanabi, S. M. & Martinez, I. Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Res.* **25**, 4692–4693 (1997).
30. Soderberg, K. A., Linehan, M. M., Ruddie, N. H. & Iwasaki, A. MAdCAM-1 expressing sacral lymph node in the lymphotoxin β -deficient mouse provides a site for immune generation following vaginal herpes simplex virus-2 infection. *J. Immunol.* **173**, 1908–1913 (2004).

A bimodular mechanism of calcium control in eukaryotes

Henning Tidow^{1,2*}, Lisbeth R. Poulsen^{1,3*}, Antonina Andreeva⁴, Michael Knudsen^{1,5}, Kim L. Hein^{1,2†}, Carsten Wiuf⁶, Michael G. Palmgren^{1,3} & Poul Nissen^{1,2}

Calcium ions (Ca^{2+}) have an important role as secondary messengers in numerous signal transduction processes^{1–4}, and cells invest much energy in controlling and maintaining a steep gradient between intracellular (~ 0.1 -micromolar) and extracellular (~ 2 -millimolar) Ca^{2+} concentrations¹. Calmodulin-stimulated calcium pumps, which include the plasma-membrane Ca^{2+} -ATPases (PMCA), are key regulators of intracellular Ca^{2+} in eukaryotes^{5–8}. They contain a unique amino- or carboxy-terminal regulatory domain responsible for autoinhibition, and binding of calcium-loaded calmodulin to this domain releases autoinhibition and activates the pump. However, the structural basis for the activation mechanism is unknown and a key remaining question is how calmodulin-mediated PMCA regulation can cover both basal Ca^{2+} levels in the nanomolar range as well as micromolar-range Ca^{2+} transients generated by cell stimulation⁷. Here we present an integrated study combining the determination of the high-resolution crystal structure of a PMCA regulatory-domain/calmodulin complex with *in vivo* characterization and biochemical, biophysical and bioinformatics data that provide mechanistic insights into a two-step PMCA activation mechanism mediated by calcium-loaded calmodulin. The structure shows the entire PMCA regulatory domain and reveals an unexpected 2:1 stoichiometry with two calcium-loaded calmodulin molecules binding to different sites on a long helix. A multifaceted characterization of the role of both sites leads to a general structural model for calmodulin-mediated regulation of PMCA that allows stringent, highly responsive control of intracellular calcium in eukaryotes, making it possible to maintain a stable, basal level at a threshold Ca^{2+} concentration, where steep activation occurs.

Compared with other P-type ATPases, calmodulin-stimulated calcium pumps contain an additional autoinhibitory or regulatory (R) domain at the N terminus⁹ (in plants) or C terminus^{10–12} (in mammals) (Supplementary Fig. 1). Pump activity is stimulated by binding of calcium-loaded calmodulin (Ca^{2+} -CaM) to the regulatory domain, thereby inducing a conformational change that displaces the autoinhibitory domain from the pump core structure^{13,14}. Calmodulin (CaM) substantially increases both Ca^{2+} affinity and pump rate in PMCA¹. To investigate the mechanism of CaM-mediated PMCA regulation, we have determined the structure of an intact regulatory domain (residues 40–95) of the *Arabidopsis thaliana* PMCA (we denote this domain Aca8R) in a homologous complex with *A. thaliana* Cam7 and characterized it *in vivo* and *in vitro*.

The crystal structure of this complex, determined at a resolution of 1.95 Å, shows an overall dumbbell shape and reveals an unexpected 2:1 stoichiometry (Fig. 1a and Supplementary Table 1): Aca8R forms a 56-residue-long α -helix with two Ca^{2+} -Cam7 molecules bound at two distinct CaM-binding sites within the autoinhibitory domain in a

pseudo-symmetrically ‘inverted’ arrangement. Given this structure, we denote the complex (Cam7)₂-Aca8R. The two binding sites are separated in space by eight residues with no interaction between the two CaM molecules. Although the first binding site (CaMBS1) had been previously mapped by mutational studies¹⁵, the second binding site (CaMBS2) was unknown and unexpected.

Both sites show typical structural characteristics of CaM-target recognition¹⁶. The hydrophobic anchor points are Trp 47 and Phe 60 in CaMBS1 and Ile 79 and Phe 92 in CaMBS2. Each anchor residue is docked in a hydrophobic cage formed in each lobe of Cam7 (Fig. 1b–d and Supplementary Figs 2 and 3). In both of the binding sites the anchor points show a 1–14 spacing, whereas in CaM complexes of human PMCA4b¹⁷ and plant BCA1¹⁸ a 1–18 spacing of anchor residues has been observed. A superimposition of the two bound CaM molecules reveals that the molecule bound to CaMBS1 has a slightly more closed conformation. The long α -helix formed by the regulatory domain of Aca8 shows a minor bending in CaMBS1.

Small-angle X-ray scattering (SAXS) data of the purified Cam7–Aca8R complex in solution agree with the crystal structure of the 2:1 complex. The distance distribution function shows two separated peaks indicating the presence of two subdomains (Supplementary Fig. 4), and the *ab initio* SAXS model superimposes with the crystal structure (Supplementary Fig. 4). In contrast, a putative 1:1 Cam7–Aca8R binding model does not fit the experimental SAXS data, indicating that the 2:1 (Cam7)₂-Aca8R complex is indeed the favoured, stable complex in solution.

Using isothermal titration calorimetry (ITC), we measured CaM binding to the isolated binding sites as well as to a construct containing the entire autoinhibitory/regulatory domain of the protein Aca8. Cam7 binds tightly to peptides corresponding to CaMBS1 (dissociation constant, $K_d = 13$ nM) and CaMBS2 ($K_d = 0.5$ μ M) of Aca8 (Fig. 1e, f and Supplementary Table 2). When measuring CaM binding to a construct comprising the entire autoinhibitory/regulatory domain (Aca8(40–126)), biphasic binding to two CaM-binding sites was observed with affinities similar to those of binding to individual peptides (Fig. 1g), whereas apo-CaM (stripped using EDTA) did not bind to any construct (data not shown). This indicates that the regulatory domain of Aca8 indeed contains two independent sites that bind Ca^{2+} -CaM with different, but physiologically relevant, affinities.

We performed functional complementation assays using a Ca^{2+} -ATPase-deficient yeast strain (K616^{19,20}). This allowed us to investigate the autoinhibitory function of the R domain from PMCA gain-of-function mutations, because heterologous expression of a constitutively active Ca^{2+} pump in the endoplasmic reticulum of the yeast is required to maintain internal calcium stores on growth in EGTA-containing medium (see Supplementary Information for a detailed account of this model). Indeed, without galactose-driven expression

¹Centre for Membrane Pumps in Cells and Disease — PUMPKIN, Aarhus University, Gustav Wieds Vej 10c, DK-8000 Aarhus C, Denmark. ²Department of Molecular Biology and Genetics, Aarhus University, Gustav Wieds Vej 10c, DK-8000 Aarhus C, Denmark. ³Department of Plant Biology and Biotechnology, University of Copenhagen, Thorvaldsensvej 40, DK-1871 Frederiksberg C, Denmark. ⁴MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, UK. ⁵Bioinformatics Research Centre, Aarhus University, CF Møllers Allé 8, DK-8000 Aarhus C, Denmark. ⁶Department of Mathematical Sciences, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen, Denmark. †Present address: Centre for Molecular Medicine Norway, Nordic EMBL Partnership, University of Oslo, PO Box 1125, Blindern, N-0318 Oslo, Norway.

*These authors contributed equally to this work.

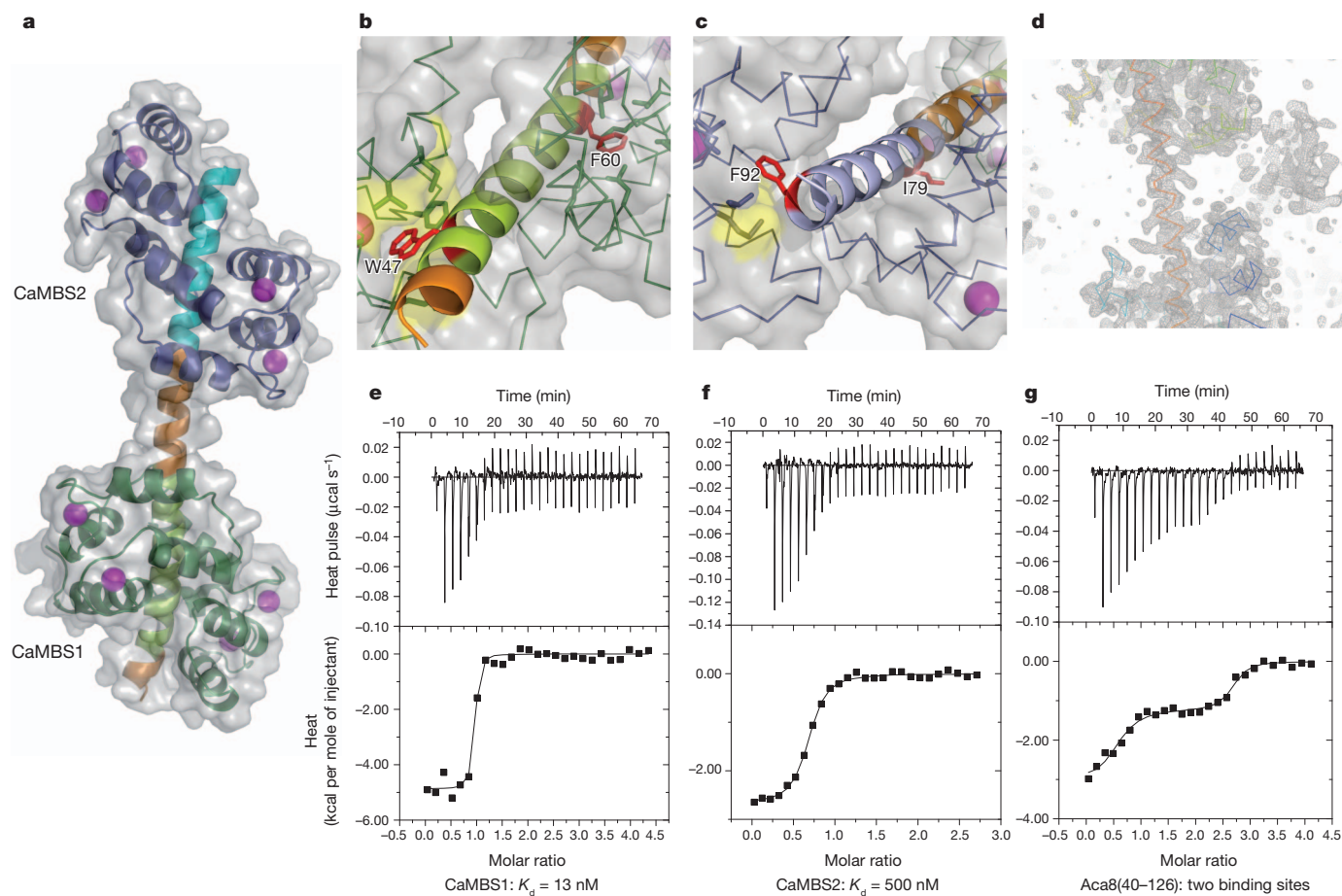


Figure 1 | Overall structure of the *A. thaliana* (Cam7)₂-Aca8R complex.

a, Representation with CaM molecules in dark green (CaMBS1) and dark blue (CaMBS2), and Aca8R in orange, light green (CaMBS1) and cyan (CaMBS2). Ca²⁺ is shown in magenta. **b, c**, Detailed view of hydrophobic anchor residues involved in CaM-binding in CaMBS1 (**b**) and CaMBS2 (**c**). Cam7 is shown as ribbon. Residues in Cam7 that are within van der Waals bonding distance of the Aca8R anchor residues (red sticks) are shown as sticks with yellow surfaces. **d**, Electron density illustrating the extended α -helix (containing both

CaM-binding sites) formed by the autoinhibitory region. **e–g**, Binding of CaM to the autoinhibitory domain of Aca8, examined by ITC. Typical ITC data are shown for the binding of Cam7 to 24-mer peptides corresponding to CaMBS1 (**e**) and CaMBS2 (**f**) of Aca8 and to Aca8(40–126) containing both CaM-binding sites (**g**). Upper panels: raw heat pulse data. Lower panels: integrated heat pulses, normalized per mole of injectant, giving differential binding curves that are adequately described by a one-site (**e** and **f**) or two-site (**g**) binding model with $K_d = 13$ nM for CaMBS1 and $K_d = 500$ nM for CaMBS2.

of Aca8 variants, none of the K616 yeast grows in Ca²⁺-depleted medium (Fig. 2a, SD medium). However, expression of a constitutively active Ca²⁺ pump by virtue of truncation of either one or both CaM-binding sites ($\Delta 74$ Aca8 or $\Delta 100$ Aca8) complements the lack of endogenous yeast Ca²⁺-ATPases and allows growth in the Ca²⁺-depleted medium (Fig. 2a, SG medium), whereas the wild-type autoinhibited Aca8 does not. Yeasts with anchor-point mutations in Aca8 CaMBS1 (Trp47Ala and Phe60Ala) but not in CaMBS2 (Phe92Ala) show moderate growth, indicating that displacement of CaMBS1 alone generates basal pump activity. The Trp47Ala Phe92Ala double mutant with mutation in both CaM-binding sites shows significantly better growth (Fig. 2a) than the Trp47Ala mutant, indicating further involvement of anchor point Phe92Ala (CaMBS2) in autoinhibition. Expression levels were similar for all Aca8 constructs (Supplementary Fig. 5). The results confirm that both CaM-binding sites contribute to autoinhibition of the pump.

The Ca²⁺-stimulated ATPase activity of wild-type Aca8 further shows that the pump encompasses both a high- and a low-affinity sensor for Ca²⁺ (via Ca²⁺-CaM; Fig. 2b). Wild-type Aca8 is inactive (autoinhibited) in the absence of CaM and active in its presence. The $\Delta 74$ Aca8 truncated form, lacking the high-affinity CaM-binding site, shows basal activity and can be further activated twofold by CaM, whereas the $\Delta 100$ Aca8 form, which lacks both CaM-binding sites, shows CaM-independent activity (Fig. 2b). Furthermore, the characterization

of the different Aca8 mutants supports the proposal that both CaM-binding sites contribute to autoinhibition as well as to Ca²⁺-CaM-mediated activation of the pump (Fig. 2c).

The presence of two CaM-binding sites is, however, not unique to the *A. thaliana* Aca8 protein. Sequence analysis suggests that both CaM-binding sites are conserved in other *Arabidopsis* Aca isoforms and plant PMCA homologues (Supplementary Fig. 6). Even more surprisingly, our analysis of mammalian PMCA (PMCA1 to PMCA4) sequences indicates that tandem CaM-binding sites are generated through alternative splicing in several variants (Fig. 3). The autoinhibitory domain of mammalian PMCA is C-terminal and subject to tissue- and cell-type-specific alternative splicing^{5,21}. PMCA splice variants show not only differences with respect to inhibition and CaM activation but also in response to different Ca²⁺ signals⁵, supporting the idea that alternative splicing might be a general mechanism for the fine-tuning of intracellular signalling^{22–24}. Binding experiments by ITC using peptides of PMCA splice variants (PMCA1 to PMCA4) confirmed that CaM binds to their second sites with K_d values in the low-micromolar range (Fig. 3a–d and Supplementary Table 2), indicating that several mammalian PMCA variants also interact with two CaM molecules. The spacing between the CaM-binding sites is comparable to that in Aca8, suggesting a similar 2:1 structural arrangement and a similar mechanism of two-step, Ca²⁺-dependent activation by CaM.

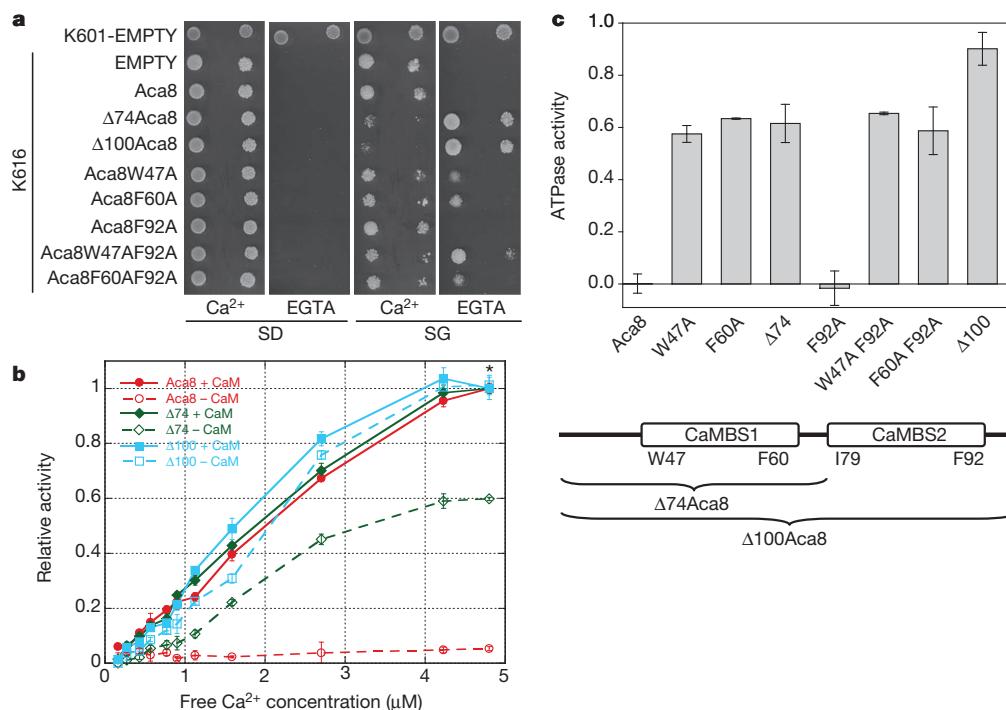


Figure 2 | Physiological and biochemical characterization of wild-type Aca8 and mutants. **a**, Functional complementation of the Ca²⁺-ATPase-deficient yeast strain (K616) expressing wild-type and mutants of Aca8. Yeast strains K601 (wild type) and K616 transformed with empty vector are used as positive and negative controls, respectively. K616 transformed with wild-type and mutants of Aca8 regulated by a galactose-inducible promoter are grown on galactose (SG) or glucose (SD) medium supplemented with either 10 mM CaCl₂ (Ca²⁺) or 10 mM EGTA. Only deregulated pumps are able to complement the lack of endogenous yeast Ca²⁺-ATPases in Ca²⁺-depleted medium (SG EGTA). **b**, Ca²⁺-stimulated ATPase activity of wild-type Aca8 and truncation variants lacking either CaMBS1 ($\Delta 74$ Aca8) or both CaM-binding sites ($\Delta 100$ Aca8) in the presence or absence of

10 μ M Cam7. The asterisk denotes the concentration of free Ca²⁺ measured by the Ca²⁺-sensitive dye fura-2 (4.80 μ M), which gave rise to the greatest difference between the activities with and without CaM and was used for normalization (each construct was normalized separately to allow direct comparison of the effect of CaM) and further characterization. The plot shows the results of one of two independent experiments carried out in triplicate. Errors, s.d. **c**, Top: ATPase activity of wild-type and mutant Aca8. The ratio of CaM-independent activity (Ca²⁺ with no Cam7) is given relative to the Ca²⁺-CaM-dependent activity (Ca²⁺ with 10 μ M Cam7). Errors, s.d. (2–4 independent experiments in triplicate). Bottom: schematic of the two CaM-binding sites; the position of anchor point residues and the parts removed in the truncated versions of Aca8 are indicated.

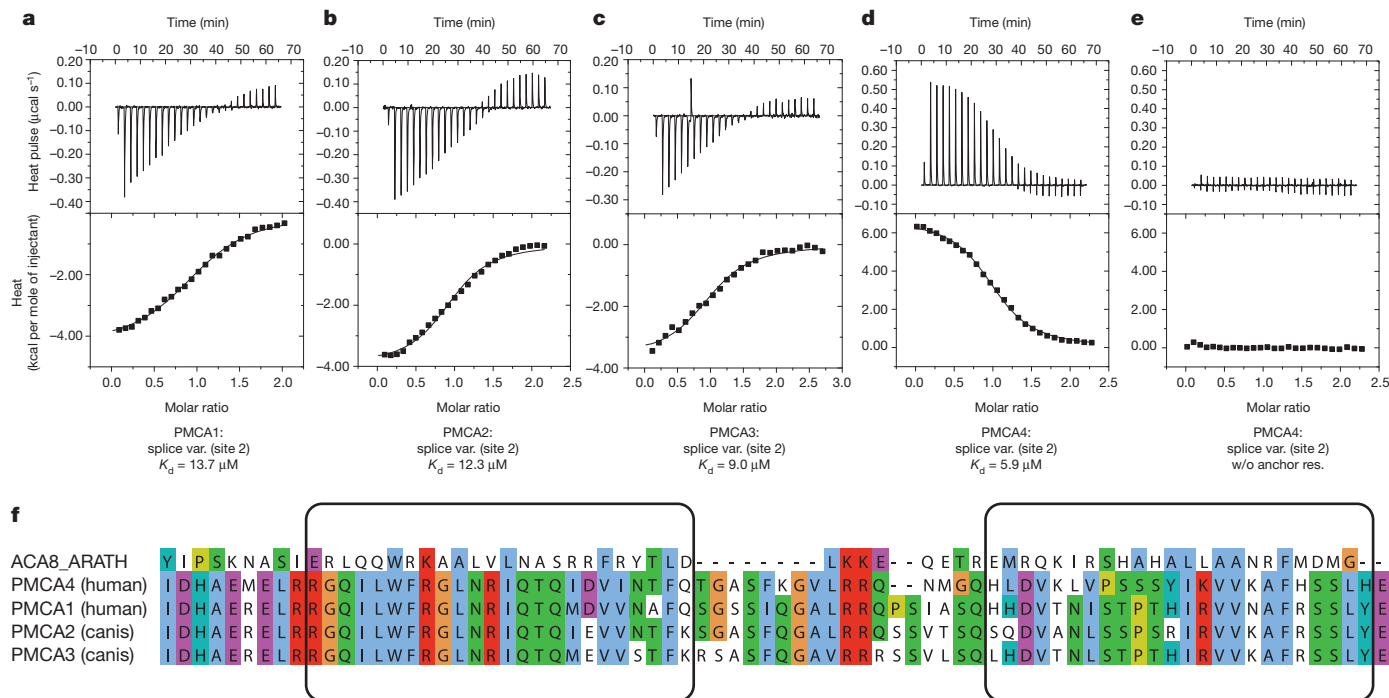


Figure 3 | Binding of CaM to the putative second CaM-binding site of PMCA splice variants. **a–d**, Typical ITC data for the binding of mammalian CaM to 24-mer peptides corresponding to splice variants of PMCA1 (**a**), PMCA2 (**b**), PMCA3 (**c**), and PMCA4 (**d**). Upper panels: raw heat pulse data. Lower panels: integrated heat pulses, normalized per mole of injectant, giving differential

binding curves that are adequately described by a one-site binding model. **e**, Control peptide lacking the hydrophobic anchor residues of PMCA4d (Val4Ala, Leu6Ala and Phe19Ala), illustrating that CaM binding is specific. **f**, Alignment of Aca8 and mammalian PMCA splice variants coloured according to the CLUSTAL colouring scheme. The two CaM-binding sites are enclosed with boxes.

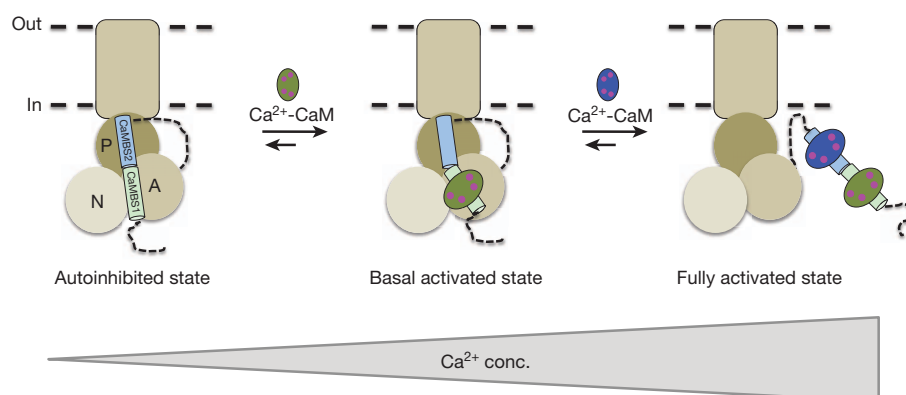


Figure 4 | Schematic of the proposed two-step, Ca^{2+} -mediated CaM-activation mechanism. With increasing Ca^{2+} concentration, Ca^{2+} -CaM first binds to and displaces high-affinity CaMBS1 before even higher Ca^{2+} concentration leads to displacement of CaMBS2 from the catalytic core,

Using mathematical network modelling of two-binding-site Ca^{2+} -CaM activation of PMCA, we find that the pump system is ready for steep activation above a basal Ca^{2+} concentration, meaning that the enzyme is inactive below this level and rapidly activates as soon as concentrations exceed it (Supplementary Fig. 7). This gives our results particular physiological relevance, because Ca^{2+} activation through just one or no CaM-binding or autoinhibitory site (assuming otherwise identical conditions) is much less abrupt, meaning that a minimal cellular Ca^{2+} concentration would be less stable and acceleration of pump activity at increased Ca^{2+} levels less pronounced.

By combining structural, physiological and biochemical data with modelling, we also derive a structural model for Aca8 in its autoinhibited state. Aca8 homology models in Ca^{2+} -bound (E1) and Ca^{2+} -free (E2) conformations reveal the presence of a highly conserved cleft, situated between the A domain and the N and P domains, that is fully exposed in the E2 conformation and buried in the E1 conformation (Supplementary Video 1 and Supplementary Figs 8 and 9). Several acidic residues previously reported to be involved in autoinhibition (Glu 252, Asp 303 and Asp 332) cluster along this region²⁵ (Supplementary Fig. 9). The N-terminal autoinhibitory domain of Aca8, however, contains several conserved basic residues¹⁵ (Arg 58, Arg 61, Lys 67 and Lys 68), which are of critical importance for autoinhibition and could potentially interact with the acidic residues mentioned above (Supplementary Fig. 8). We suggest that as the Ca^{2+} concentration increases, Ca^{2+} -CaM first binds to and displaces the high-affinity CaMBS1 in the region between the A and N domains, and that further Ca^{2+} -CaM activation leads to displacement of CaMBS2 from the linker regions between the transmembrane domain and the mobile A domain (Fig. 4 and Supplementary Fig. 8). Displacement of CaMBS1 allows free movement of the cytoplasmic domains as required for ion pumping at basal levels, and the additional displacement of CaMBS2 by higher Ca^{2+} -CaM concentrations or the presence of acidic phospholipids²⁶ results in further activity. This model provides a conceptual framework for a bimodular, Ca^{2+} -mediated CaM-activation mechanism that allows regulation of intracellular Ca^{2+} concentration over a broad range of physiological conditions, ready to respond promptly at increased Ca^{2+} levels and setting the basal Ca^{2+} level in the cell. Such a mechanism seems to be well suited to the fine-tuning of calcium homeostasis and intracellular signalling in eukaryotes.

METHODS SUMMARY

Procedures for the expression, purification and crystallization of the (Cam7)₂-Aca8R complex were similar to those described in ref. 27. The complex was crystallized at 4 °C using sitting-drop vapour diffusion against a reservoir containing 1.9 M $(\text{NH}_4)_2\text{SO}_4$, 0.1 M CAPS (pH 10.5) and 0.2 M Li_2SO_4 . The crystals have space group $P4_12_12$ and unit-cell parameters $a = b = 71.25$ Å, $c = 163.28$ Å and

allowing free movement of the A domain as required for ion pumping. Actuator (A), nucleotide-binding (N) and phosphorylation (P) domains and the transmembrane region are indicated. Aca8R is shown as cylinder with the CaM-binding site colour-coded as in Fig. 1.

$\alpha = \beta = \gamma = 90^\circ$. Diffraction data were collected to a resolution limit of 1.95 Å at the ID23-2 beamline of the European Synchrotron Radiation Facility (France). The structure was determined by molecular replacement followed by iterative model building and refinement. The final model yielded a crystallographic R -factor of 22.1% and a free R -factor of 25.3%. Evaluation of the Ramachandran plot showed all residues except one (99.7%) in allowed regions (97.0% in favoured regions). SAXS data were collected at the X33 beamline at EMBL/DESY (Germany), following standard procedures. Heterologous expression, yeast complementation tests, membrane purification and activity assays of full-length Aca8, N-terminal truncation and single-point-mutation constructs were performed essentially as previously described^{15,28}. More detailed methods can be found in Method and Supplementary Information.

Full Methods and any associated references are available in the online version of the paper.

Received 2 December 2011; accepted 23 August 2012.

Published online 21 October 2012.

- Clapham, D. E. Calcium signaling. *Cell* **131**, 1047–1058 (2007).
- Carafoli, E. Calcium signaling: a tale for all seasons. *Proc. Natl Acad. Sci. USA* **99**, 1115–1122 (2002).
- Ghosh, A. & Greenberg, M. E. Calcium signaling in neurons: molecular mechanisms and cellular consequences. *Science* **268**, 239–247 (1995).
- Berridge, M. J. Inositol trisphosphate and calcium signalling. *Nature* **361**, 315–325 (1993).
- Strehler, E. E., Filoteo, A. G., Penniston, J. T. & Caride, A. J. Plasma-membrane Ca^{2+} pumps: structural diversity as the basis for functional versatility. *Biochem. Soc. Trans.* **35**, 919–922 (2007).
- Brini, M. & Carafoli, E. Calcium pumps in health and disease. *Physiol. Rev.* **89**, 1341–1378 (2009).
- Brini, M. & Carafoli, E. The plasma membrane Ca^{2+} ATPase and the plasma membrane sodium calcium exchanger cooperate in the regulation of cell calcium. *Cold Spring Harb. Perspect. Biol.* **3**, 1–15 (2011).
- Axelsen, K. B. & Palmgren, M. G. Evolution of substrate specificities in the P-type ATPase superfamily. *J. Mol. Evol.* **46**, 84–101 (1998).
- Malmström, S., Askerlund, P. & Palmgren, M. G. A calmodulin-stimulated Ca^{2+} -ATPase from plant vacuolar membranes with a putative regulatory domain at its N-terminus. *FEBS Lett.* **400**, 324–328 (1997).
- Falchetto, R., Vorherr, T., Brunner, J. & Carafoli, E. The plasma membrane Ca^{2+} pump contains a site that interacts with its calmodulin-binding domain. *J. Biol. Chem.* **266**, 2930–2936 (1991).
- Falchetto, R., Vorherr, T. & Carafoli, E. The calmodulin-binding site of the plasma membrane Ca^{2+} pump interacts with the transduction domain of the enzyme. *Protein Sci.* **1**, 1613–1621 (1992).
- Bonza, M. C. & Luoni, L. Plant and animal type 2B Ca^{2+} -ATPases: evidence for a common auto-inhibitory mechanism. *FEBS Lett.* **584**, 4783–4788 (2010).
- Carafoli, E. Biogenesis: plasma membrane calcium ATPase: 15 years of work on the purified enzyme. *FASEB J.* **8**, 993–1002 (1994).
- Penniston, J. T. & Enyedi, A. Modulation of the plasma membrane Ca^{2+} pump. *J. Membr. Biol.* **165**, 101–109 (1998).
- Baekgaard, L., Luoni, L., De Michelis, M. I. & Palmgren, M. G. The plant plasma membrane Ca^{2+} pump ACA8 contains overlapping as well as physically separated autoinhibitory and calmodulin-binding domains. *J. Biol. Chem.* **281**, 1058–1065 (2006).
- Hoeflich, K. P. & Ikura, M. Calmodulin in action: diversity in target recognition and activation mechanisms. *Cell* **108**, 739–742 (2002).

17. Juranic, N. *et al.* Calmodulin wraps around its binding domain in the plasma membrane Ca^{2+} pump anchored by a novel 18–1 motif. *J. Biol. Chem.* **285**, 4015–4024 (2010).
18. Ishida, H. & Vogel, H. J. The solution structure of a plant calmodulin and the CaM-binding domain of the vacuolar calcium-ATPase BCA1 reveals a new binding and activation mechanism. *J. Biol. Chem.* **285**, 38502–38510 (2010).
19. Cunningham, K. W. & Fink, G. R. Calcineurin-dependent growth control in *Saccharomyces cerevisiae* mutants lacking PMC1, a homolog of plasma membrane Ca^{2+} ATPases. *J. Cell Biol.* **124**, 351–363 (1994).
20. Durr, G. *et al.* The medial-Golgi ion pump Pmr1 supplies the yeast secretory pathway with Ca^{2+} and Mn^{2+} required for glycosylation, sorting, and endoplasmic reticulum-associated protein degradation. *Mol. Biol. Cell* **9**, 1149–1162 (1998).
21. Strehler, E. E. & Zacharias, D. A. Role of alternative splicing in generating isoform diversity among plasma membrane calcium pumps. *Physiol. Rev.* **81**, 21–50 (2001).
22. Enyedi, A. *et al.* The Ca^{2+} affinity of the plasma membrane Ca^{2+} pump is controlled by alternative splicing. *J. Biol. Chem.* **269**, 41–43 (1994).
23. Namba, T. *et al.* Alternative splicing of C-terminal tail of prostaglandin E receptor subtype EP3 determines G-protein specificity. *Nature* **365**, 166–170 (1993).
24. Spengler, D. *et al.* Differential signal transduction by five splice variants of the PACAP receptor. *Nature* **365**, 170–175 (1993).
25. Fusca, T. *et al.* Single point mutations in the small cytoplasmic loop of ACA8, a plasma membrane Ca^{2+} -ATPase of *Arabidopsis thaliana*, generate partially deregulated pumps. *J. Biol. Chem.* **284**, 30881–30888 (2009).
26. Meneghelli, S., Fusca, T., Luoni, L. & De Michelis, M. I. Dual mechanism of activation of plant plasma membrane Ca^{2+} -ATPase by acidic phospholipids: Evidence for a phospholipid binding site which overlaps with the calmodulin-binding site. *Mol. Membr. Biol.* **25**, 539–546 (2008).
27. Tidow, H., Hein, K. L., Baekgaard, L., Palmgren, M. G. & Nissen, P. Expression, purification, crystallization and preliminary X-ray analysis of calmodulin in complex with the regulatory domain of the plasma-membrane Ca^{2+} -ATPase ACA8. *Acta Crystallogr. F* **66**, 361–363 (2010).
28. Schiott, M. *et al.* A plant plasma membrane Ca^{2+} pump is required for normal pollen tube growth and fertilization. *Proc. Natl Acad. Sci. USA* **101**, 9502–9507 (2004).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank members of the Nissen and Palmgren labs for discussions, and P. Gourdon for help with data collection. We thank the staff at beamlines ID23-2 at the European Radiation Synchrotron Facility, France; PX3 at the Swiss Light Source, Paul Scherrer Institute, Switzerland; and X33 at EMBL/DESY, Germany. We are grateful to K. Nagai for a plasmid expressing mammalian CaM. Support from the European Community-Research Infrastructure Action under the FP7 is acknowledged for access to EMBL/DESY. H.T. is a Junior Research Fellow at Trinity College, Cambridge, and was supported by an EMBO Long-Term Fellowship, a Marie-Curie Intra-European Fellowship and an HFSP Long-Term Fellowship. P.N. was supported by an ERC advanced grant (BIOMEMOS).

Author Contributions H.T. designed and initiated the project, designed the expression constructs and developed the co-expression strategy, initially assisted by K.L.H. Protein purification, crystallization, structure determination and refinement, and the overall analysis of the results, was performed by H.T. L.R.P. performed biochemical and genetic analyses of Aca8 and derived mutants, developed methods for measuring calcium concentrations *in vitro*, and analysed biochemical and yeast complementation assays, supervised by M.G.P. A.A. performed bioinformatics sequence analysis, homology modelling and docking experiments. M.K. performed mathematical modelling, supervised by C.W. P.N. designed and supervised the project, and analysed results. H.T., L.R.P., A.A., M.G.P. and P.N. wrote the paper, and all authors commented on the paper.

Author Information Atomic coordinates and structure factors for the crystal structure of (Cam7)₂-Aca8R have been deposited with the Protein Data Bank under accession code 4AQR. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.G.P. (palmgren@life.ku.dk) or P.N. (pn@mb.au.dk).

METHODS

Purification and crystallization of (Cam7)₂-Aca8R. Purification of the complex and initial crystallization was performed as previously described²⁷. In brief, Aca8(40–95) and Cam7 (both from *A. thaliana*) were co-expressed in *Escherichia coli* C41 cells²⁹ and purified using standard His-tag purification protocols followed by TEV protease digestion, a second Ni-affinity chromatography step to separate a fusion protein tag and a final gel-filtration step. Initial crystals were obtained using the vapour diffusion technique in sitting drops prepared by mixing 1.5 µl of (Cam7)₂-Aca8R solution (16 mg ml⁻¹) with 1 µl of reservoir solution, and equilibrated against reservoir solution containing 2.0 M (NH₄)₂SO₄, 0.1 M CAPS (pH 10.5) and 0.2 M Li₂SO₄ (final pH 8.2) at room temperature (293 K). These crystals belonged to space group C2 and diffracted to a resolution of 3.0 Å (ref. 27). Better-diffracting crystals belonging to a different space group (P4₁2₁2) were obtained at 4 °C under slightly modified conditions: 1 µl of (Cam7)₂-Aca8R solution (16 mg ml⁻¹) was mixed with 1 µl of reservoir solution and equilibrated against reservoir solution containing 1.9 M (NH₄)₂SO₄, 0.1 M CAPS (pH 10.5) and 0.2 M Li₂SO₄ at 4 °C (277 K). Crystals appeared after several months, grew to maximum dimensions of 0.7 mm × 0.35 mm × 0.2 mm within weeks (Supplementary Fig. 10) and proved to be superior to the room-temperature crystal with respect to diffraction properties. Crystals were mounted and flash-cooled in liquid nitrogen without additional cryoprotection.

Characterization of CaM-binding sites using complementation tests and activity assays. Cam7 from *A. thaliana* and mammalian CaM alone were expressed in *E. coli* C41 and purified using standard His-tag purification protocols. Heterologous expression, yeast complementation tests, membrane purification and activity assays of full-length Aca8, N-terminal truncation and single-point-mutation constructs were performed essentially as previously described^{15,28}. The yeast strain K616 lacks the regulatory subunit B (Cnb1) of calcineurin in addition to the two Ca²⁺-ATPases (Pmc1 and Pmr1). Under normal conditions, calcineurin inhibits the low-affinity vacuolar Ca²⁺/H⁺ antiporter Vcx1. Vcx1 is the only transporter that can fill up the vacuolar Ca²⁺ store in the mutant lacking the two Ca²⁺-ATPases. At high Ca²⁺ concentration in the cytosol, Vcx1 transports Ca²⁺ into the vacuole. Therefore, to get a viable yeast strain, the *cnb1* mutation is required. The mutation of calcineurin activates the antiporter, which then removes Ca²⁺ from the cytosol. At low Ca²⁺ concentration, the K616 strain needs an active high-affinity transporter (Ca²⁺-ATPase) to remove Ca²⁺ from the cytosol; that is, complementation requires a constitutively active Ca²⁺ pump in the endoplasmic reticulum/Golgi apparatus apparently to scavenge trace Ca²⁺ for proper functioning of the secretory pathway. It has previously been shown that only expression of active Ca²⁺-ATPases can rescue the growth defect in Ca²⁺-depleted medium¹⁵. The growth rate of the mutant in Ca²⁺-depleted medium thus provides an *in vivo* assay for functional characterization of overexpressed recombinant Ca²⁺ pumps from heterologous sources.

The expression level of Aca8 and mutants in yeast microsomal membranes (20 µg) were determined by SDS-polyacrylamide gel electrophoresis. The biochemical characterization was carried out on microsomal membranes (5 µg). ATPase activity was determined with and without 10 µM Cam7 by varying the free Ca²⁺ concentration. The free Ca²⁺ concentrations were determined using the Ca²⁺-sensitive dye fura-2 (ref. 30). Calibration curves were generated using the Calcium Calibration Buffer Kit #1 (zero and 10 mM CaEGTA) and fura-2 pentapotassium salt (Invitrogen/Molecular Probes). Each buffer and ATPase assay sample (prepared as for the ATPase activity assay) was supplemented with 1 µM fura-2. For each calibration buffer and test sample, excitation scans were generated between wavelengths 250 and 450 nm (slit, 5 nm) while detecting the emission at wavelength 510 nm (slit, 5 nm) essentially as described in the provided protocol from Invitrogen. The HORIBA Jobin Yone FluoroMax-4 spectrofluorometer and the software FLUORESCENCE were used for the measurements. For the final calculations, fluorescence at 380 nm (F380) was used. The free Ca²⁺ concentration in the calibration kit buffer was calculated using the specific K_d (EGTA) values at pH 7.3 (30 °C). The addition of Cam7 in the absence of Ca²⁺ did not result in any change in ATPase activity. Activity without Ca²⁺ and Cam7 were subtracted from those with Ca²⁺ and with and without Cam7 to give the measured Ca²⁺-ATPase activity.

Data collection and X-ray diffraction analysis. Diffraction data were collected to a resolution limit of 1.95 Å. Two full data sets containing 200 oscillation images each with an interval of 1° were collected using the same crystal at a wavelength of 0.9464 Å at the ID23-2 beamline of the European Synchrotron Radiation Facility (France). Exposure times and crystal-to-detector distances were respectively 0.4 s and 472 mm (set one) and 0.5 s and 245 mm (set two). Low-resolution and high-resolution data sets were merged using XSCALE³¹.

Reflections were indexed and scaled with XDS³¹. The crystals belonged to the tetragonal space group P4₁2₁2 and had unit-cell parameters *a* = *b* = 71.25 Å,

c = 163.28 Å and $\alpha = \beta = \gamma = 90^\circ$. A summary of the data statistics is given in Supplementary Table 1.

Structure determination. The structure was solved by molecular replacement using PHASER³² with a systematic combination of search models (derived from various CaM complexes) and search parameters. An initial molecular replacement solution contained two molecules of CaM N-terminal domain and one molecule of CaM C-terminal domain and revealed clear extra density that could be unambiguously assigned to Aca8R. Subsequent rounds of manual building using COOT³³ and refinement using PHENIX.REFINE³⁴ allowed placement of the second CaM C-terminal domain and led to almost complete model building (five residues (76–80) in a flexible linker connecting Cam7 lobes bound to CaMBS2 could not be traced). The final model yielded a crystallographic *R*-factor of 22.1% and a free *R*-factor of 25.3%. The model was validated using MOLPROBITY^{35,36}. Evaluation of the Ramachandran plot showed all residues except one (99.7%) in allowed regions (97.0% in favoured regions), with Ser 82 situated in the flexible loop connecting the N- and C-terminal lobes of Cam7 (CaMBS1) as the single outlier. All figures were prepared using PYMOL³⁷. The data are deposited in the Protein Data Bank.

SAXS — data collection and *ab initio* modelling. SAXS data were collected at the X33 beamline at EMBL/DESY (Germany), following standard procedures. Repetitive data collection on the same sample was performed and no radiation damage was detected. Samples of (Cam7)₂-Aca8R solution were prepared in concentrations ranging from 4.5 to 9.0 mg ml⁻¹ in a buffer containing 40 mM Tris/HCl (pH 7.4), 150 mM NaCl, 5 mM β-mercaptoethanol and 5 mM CaCl₂. All SAXS data were analysed using the package ATSAS³⁸. Raw data were processed using PRIMUS³⁹. The radius of gyration (*R*_g) was evaluated using the Guinier approximation ($I(s) = I(0)\exp(-s^2 R_g^2/3)$ for $sR_g < 1.3$, where *I* denotes intensity and *s* denotes momentum transfer) and also from the entire scattering curve with the program GNOM⁴⁰; the latter also provided the distance distribution function, *p*(*r*), and the maximum dimension, *D*_{max}. The masses of the solutes were evaluated by comparison of the forward scattering intensity with that from a BSA reference solution (mass, 66 kDa).

Low-resolution SAXS models were obtained using the *ab initio* simulated annealing program DAMMIN⁴¹, which generates models consisting of dummy atoms to fit the experimental data *I*_{exp}(*s*) by minimizing the discrepancy

$$\chi^2 = \frac{1}{N-1} \sum_j \left(\frac{I_{\text{exp}}(s_j) - c I_{\text{calc}}(s_j)}{\sigma(s_j)} \right)^2$$

where *N* is the number of experimental points, *c* is a scaling factor, and *I*_{calc}(*s*_{*j*}) and $\sigma(s_j)$ are the calculated intensity and the experimental error at the momentum transfer *s*_{*j*}, respectively. Superimposition of low-resolution dummy atom models was performed using SUPCOMB⁴².

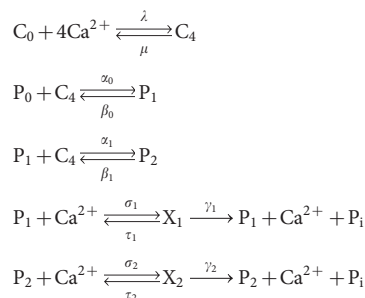
Isothermal titration calorimetry. The thermodynamic parameters were determined using an isothermal titration calorimeter (MicroCal ITC₂₀₀, GE Healthcare) at 15 °C in a buffer containing 40 mM Tris/HCl (pH 7.4), 150 mM NaCl, 5 mM β-mercaptoethanol, 5 mM CaCl₂. For binding of Cam7 to Aca8(40–126) 400 µM Cam7 was titrated into the sample cell containing 20 µM Aca8(40–126). For binding of Cam7 to peptides corresponding to Aca8 CaMBS1 (residues 42–65) and Aca8 CaMBS2 (residues 74–97) typically 150–400 µM of the peptide was titrated into the sample cell containing 7–40 µM Cam7. Reverse titration yielded identical thermodynamic parameters. For binding of mammalian CaM to peptides corresponding to the putative second CaM-binding site of mammalian PMCA1, PMCA2, PMCA3 and PMCA4 splice variants typically 0.6–1.2 mM CaM was titrated into the sample cell containing 50–100 µM peptide. Injection steps were 1.5 µl (first injection, 0.3 µl) with 150-s spacing. Baseline corrections were performed by titrating protein or peptide into sample buffer and sample buffer into protein or peptide. Further data evaluation (Supplementary Table 2) was done using the MicroCal ORIGIN program. Peptides were purchased from GL Biochem (China; HPLC purified, >90% purity).

The peptide sequences were as follows: Aca8-CaMBS1: ERLQQWRKA ALVLNASRRFRYTLD; Aca8-CaMBS2: EMRQKIRSHAHALLAANRFMDMGR; PMCA1-CaMBS2: HHVDVTNISTPHTIRVNAFRSSLY; PMCA2-CaMBS2: SQD VANLSSPSRIRVVKAFRRSSLY; PMCA3-CaMBS2: LHDVTNLSPTHTIRVVKAFRRSSLY; PMCA4-CaMBS2: HLDVKLVPSSSYIKVVKAFHSSSLH; PMCA4-CaMBS2-minus-anchor: HLDKAVPSSYIKVVKAAHSSSLH.

Bioinformatics analysis and homology modelling. Multiple sequence alignments were produced using CLUSTAL W⁴³. A sequence motif was generated using WEBLOGO⁴⁴. Homology models of Aca8 in the E1 and E2 conformations were computed with MODELLER⁴⁵ using the SERCA structures (Protein Data Bank IDs, 3N5K and 3N8G) as templates and a manually optimized alignment as input. Conservation scores were calculated with CONSURF⁴⁶. Docking of the autoinhibitory N terminus onto the Aca8 homology model was performed with HADDOCK^{47,48}

using residues previously observed to be involved in autoinhibition^{15,25} (Arg 58, Arg 61, Lys 67, Lys 68, Glu 252, Asp 303 and Asp332) as restraints.

Mathematical modelling. We denote free CaM by C_0 and CaM in complex with four Ca^{2+} ions by C_4 . The autoinhibited wild-type pump is denoted P_0 , whereas P_1 and P_2 denote the pumps with one and, respectively, two CaM-binding sites occupied by C_4 (X_1 and X_2 in the Ca^{2+} -bound state). Finally, P_i denotes inorganic phosphorus. The wild-type model is represented by the following set of reactions:



Assuming mass-action kinetics, the reactions give rise to a set of polynomial differential equations. Supplementary Fig. 7 shows a typical output obtained using the reaction rates $\lambda = 10$, $\mu = 5$, $\alpha_0 = 10$, $\beta_0 = 0.1$, $\alpha_1 = 1$, $\beta_1 = 0.25$, $\sigma_1 = 1$, $\tau_1 = 0.001$, $\gamma_1 = 0.01$, $\sigma_2 = 2$, $\tau_2 = 0.001$ and $\gamma_2 = 0.1$ and with total amount of CaM equal to 10 and the total amount of PMCA equal to 1. For varying total amounts of Ca^{2+} , the steady-state concentrations are obtained from numerical simulation of the system of differential equations using MATHEMATICA. The steady-state production rate of inorganic phosphorus

$$\frac{d[P_i]}{dt} = \gamma_1[X_1] + \gamma_2[X_2]$$

where square brackets denote concentration, is used as a measure of pump activity and plotted as a function of the steady-state concentration of free calcium. The Δ CaMBS1 and Δ CaMBS1+2 models are derived from the wild-type model by leaving out P_0 and, respectively, P_0 and P_1 as well as all reactions in which the omitted pump or pumps participate. All simulations were started with $[Ca^{2+}] = Ca^{2+}_{tot}$ and $[C_0] = C_{tot}$, and with all PMCAs in the state with no CaM bound. That is, in states P_0 , P_1 and P_2 for wild-type, Δ CaMBS1 and Δ CaMBS1+2, respectively.

We note that the models are nested: all parameters in the Δ CaMBS1+2 and Δ CaMBS1 models are also parameters in the Δ CaMBS1 and, respectively, wild-type models. Simulations were made assuming that $\alpha_0 > \alpha_1$ and $\sigma_2 > \sigma_1$, meaning

that the binding affinity of CaM to CaMBS1 was higher than that of CaM to CaMBS2 and that the activity of P_2 was higher than that of P_1 , respectively.

29. Miroux, B. & Walker, J. E. Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J. Mol. Biol.* **260**, 289–298 (1996).
30. Gryniewicz, G., Poenie, M. & Tsien, R. Y. A new generation of Ca^{2+} indicators with greatly improved fluorescence properties. *J. Biol. Chem.* **260**, 3440–3450 (1985).
31. Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125–132 (2010).
32. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
33. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
34. Afonine, P. V., Grosse-Kunstleve, R. W. & Adams, P. D. The Phenix refinement framework. *CCP4 Newsletter* **42**, 8 (2005).
35. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
36. Davis, I. W. et al. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375–W383 (2007).
37. DeLano, W. L. The PyMOL Molecular Graphics System. <http://www.pymol.org> (2002).
38. Konarev, P. V., Petoukhov, M. V., Volkov, V. V. & Svergun, D. I. ATSAS 2.1, a program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* **39**, 277–286 (2006).
39. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36**, 1277–1282 (2003).
40. Svergun, D. I. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **25**, 495–503 (1992).
41. Svergun, D. I. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **76**, 2879–2886 (1999).
42. Kozin, M. B. & Svergun, D. I. Automated matching of high- and low-resolution structural models. *J. Appl. Crystallogr.* **34**, 33–41 (2001).
43. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
44. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).
45. Sali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815 (1993).
46. Glaser, F. et al. ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics* **19**, 163–164 (2003).
47. Dominguez, C., Boelens, R. & Bonvin, A. M. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* **125**, 1731–1737 (2003).
48. de Vries, S. J. et al. HADDOCK versus HADDOCK: new features and performance of HADDOCK2.0 on the CAPRI targets. *Proteins* **69**, 726–733 (2007).

Endothelial cell expression of haemoglobin α regulates nitric oxide signalling

Adam C. Straub¹, Alexander W. Lohman^{1,2}, Marie Billaud^{1,2}, Scott R. Johnstone¹, Scott T. Dwyer³, Monica Y. Lee¹, Pamela Schoppee Bortz¹, Angela K. Best¹, Linda Columbus⁴, Benjamin Gaston³ & Brant E. Isakson^{1,2}

Models of unregulated nitric oxide (NO) diffusion do not consistently account for the biochemistry of NO synthase (NOS)-dependent signalling in many cell systems^{1–3}. For example, endothelial NOS controls blood pressure, blood flow and oxygen delivery through its effect on vascular smooth muscle tone⁴, but the regulation of these processes is not adequately explained by simple NO diffusion from endothelium to smooth muscle^{3,5}. Here we report a new model for the regulation of NO signalling by demonstrating that haemoglobin (Hb) α (encoded by the *HBA1* and *HBA2* genes in humans) is expressed in human and mouse arterial endothelial cells and enriched at the myoendothelial junction, where it regulates the effects of NO on vascular reactivity. Notably, this function is unique to Hb α and is abrogated by its genetic depletion. Mechanistically, endothelial Hb α haem iron in the Fe³⁺ state permits NO signalling, and this signalling is shut off when Hb α is reduced to the Fe²⁺ state by endothelial cytochrome b5 reductase 3 (CYB5R3, also known as diaphorase 1)⁶. Genetic and pharmacological inhibition of CYB5R3 increases NO bioactivity in small arteries. These data reveal a new mechanism by which the regulation of the intracellular Hb α oxidation state controls NOS signalling in non-erythroid cells. This model may be relevant to haem-containing globins in a broad range of NOS-containing somatic cells^{7–13}.

Endothelial NOS modulates blood vessel diameter in response to both vasodilators and vasoconstrictors. For example, it is known that during arterial constriction, NO from endothelium feeds back on smooth muscle to control the magnitude of the response to a vasoconstrictor (for example, phenylephrine)^{5,14}. Phenylephrine stimulation of thoracodorsal arteries *ex vivo*—and of primary human endothelial cells (ECs) and vascular smooth muscle cells (SMCs) in the vascular cell co-culture (VCCC) model—reproduced classical NOS- and cyclic guanosine monophosphate (cGMP)-dependent changes in SMC biology (Supplementary Fig. 1a–d). However, NO did not diffuse into the extracellular space (Supplementary Fig. 1e–h), consistent with our previous work showing compartmentalized NOS signalling at the myoendothelial junction (MEJ), the crucial EC–SMC contact point in the thoracodorsal arteries and other small arteries and arterioles⁵. Therefore, we studied MEJ proteins that could contribute to local regulation of NO diffusion and biochemistry. We performed a proteomic analysis of MEJs isolated from VCCCs using the isobaric tags for relative and absolute quantification (iTRAQ) system (Supplementary Fig. 2). Surprisingly, Hb α was abundant at the MEJ (Supplementary Fig. 3). Because Hb can regulate NO diffusion and biochemistry in erythrocytes^{15,16}, we proposed that it could have a similar function at the MEJ.

First, we confirmed the proteomic data using immunoblot and immunofluorescence. We demonstrated Hb α protein expression in the VCCC model, but found no expression of Hb β (Fig. 1a). There was little Hb α expression in human ECs or SMCs grown separately, and there was no Hb α in the fibronectin or gelatin used to coat the VCCC transwells (Fig. 1a). Next, we confirmed these results in co-cultures of

different types of EC and SMC in which MEJs also expressed abundant Hb α (Supplementary Fig. 4). We then studied the MEJ distribution of Hb α *in situ*. Gold particles labelling Hb α were abundant in the MEJ of mouse thoracodorsal arteries visualized by transmission electron microscopy (TEM) (Fig. 1b). By contrast, carotid arteries (conduit arteries that have few MEJs) expressed little Hb α as observed by TEM (Fig. 1b), immunoblot (Fig. 1c) and immunofluorescence (Fig. 1d). These data were consistent in human skeletal muscle arterioles (Fig. 1d) and throughout several tissue beds (Supplementary Fig. 5). Using *en face* immunofluorescence, we found punctuate Hb α staining primarily at paracellular junctions of thoracodorsal, but not carotid, arteries, whereas little Hb β was observed (Fig. 1e). Chemical crosslinking analysis showed that the Hb α was monomeric in thoracodorsal arteries and in VCCCs (Fig. 1f). Next, we measured Hb α messenger RNA using real-time PCR (Fig. 1g), and established that ECs transfected with Hb α short interfering RNA (siRNA) had decreased protein expression at the MEJ (Supplementary Fig. 6a) and in the endothelial monolayer (Supplementary Fig. 6b). Loss of Hb α protein expression did not affect endothelial NOS (eNOS) expression in the EC monolayer (Supplementary Fig. 6b) or at the MEJ (Supplementary Fig. 7).

Transcripts for other globins including myoglobin, neuroglobin and cytoglobin were absent in ECs (Supplementary Fig. 8a–c). Only cytoglobin mRNA and protein were expressed in SMCs (Supplementary Fig. 8c, d), consistent with a previous report¹¹. In addition, we also found Hb α -stabilizing protein in the endothelium of thoracodorsal arteries and in VCCCs (Supplementary Fig. 9a, b). Taken together, these data show for the first time, to our knowledge, that arterial ECs express Hb α mRNA and protein, and are responsible for enriched Hb α expression at the MEJ.

To investigate the functional role of Hb α in ECs and its effect on eNOS signalling, we transfected ECs in isolated thoracodorsal arteries with Hb α or control siRNA. Knockdown efficiency was 70–80% (Supplementary Fig. 10). Loss of Hb α resulted in a marked loss in arterial reactivity after phenylephrine application in a single or cumulative doses (Fig. 2a, b) and increased reactivity to acetylcholine (ACh) (Fig. 2c), but there was no change in the response to 5-hydroxytryptamine (5-HT) (Supplementary Table 1). Half-maximum effective concentration (EC₅₀) and maximum effect (*E*_{max}) values are listed in Supplementary Table 2. We observed no difference in basal tone (Supplementary Fig. 11a). However, after the addition of the NOS inhibitor L-N^G-nitroarginine methyl ester (L-NAME), the effect of Hb α siRNA was comparable to control conditions for both the phenylephrine and ACh responses (Fig. 2a–c). We thus proposed that eNOS, the primary isoform in the vessel wall, may be in close spatial proximity to Hb α . We tested this hypothesis using four methods: co-localization studies by immunofluorescence (Fig. 2d, g), a proximity ligation assay (Fig. 2e), and co-immunoprecipitations from both cell lysates (Fig. 2f, h) and purified proteins (Fig. 2i). These analyses showed that Hb α

¹Robert M. Berne Cardiovascular Research Center, University of Virginia, Charlottesville, Virginia 22908, USA. ²Department of Molecular Physiology and Biological Physics, University of Virginia, Charlottesville, Virginia 22908, USA. ³Department of Pediatrics, University of Virginia, Charlottesville, Virginia 22908, USA. ⁴Department of Chemistry, University of Virginia, Charlottesville, Virginia 22908, USA.

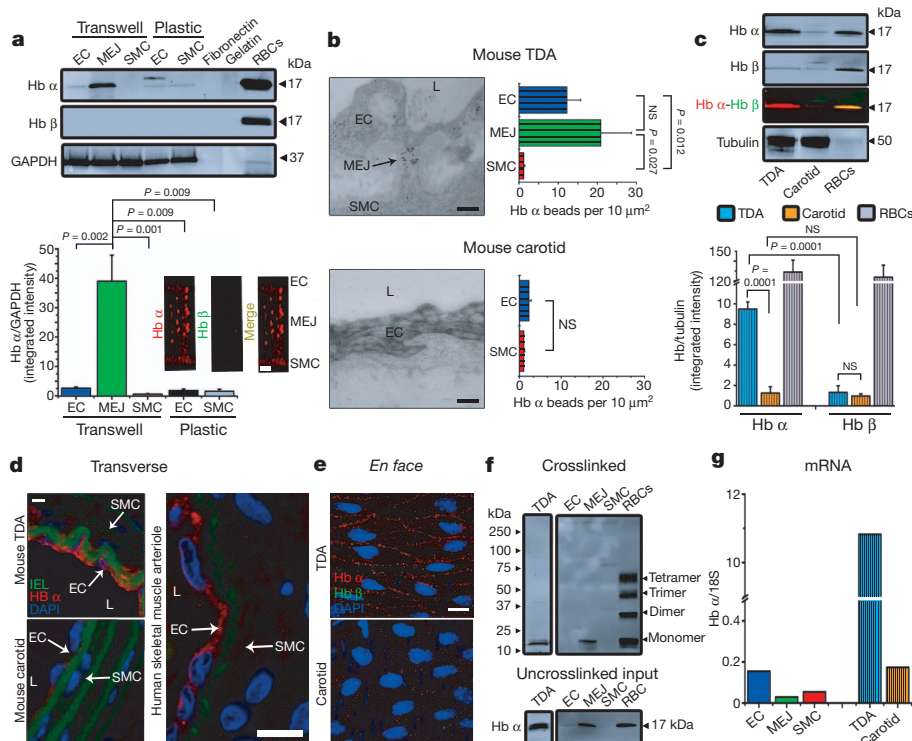


Figure 1 | Monomeric Hb α is expressed in ECs and enriched at the MEJ.

a, Quantitative western blot analysis for Hb α and Hb β expression in coronary EC, MEJ or SMC lysates from cells plated on Transwells or plastic ($n \geq 4$), or on fibronectin and gelatin used to coat Transwells. Red blood cells (RBCs) served as a positive control and GAPDH was used as a loading and normalization control for quantification (bottom left). Immunofluorescence for Hb α (red) and Hb β (green) (bottom right). Ctrl, control. **b**, TEM analysis of Hb α expression in mouse thoracodorsal arteries (TDA; top) or carotid arteries (bottom), visualized using 10-nm gold beads (black particles). Arrow indicates MEJ; L denotes lumen. Graphs on the right represent Hb α localization calculated by measuring the number of beads per $10 \mu\text{m}^2$ ($n \geq 7$). **c**, Quantitative western blot analysis for Hb α and Hb β expression in isolated thoracodorsal or carotid arteries. Tubulin served as a loading control and red blood cells were used as a positive control ($n \geq 3$). **d**, Immunofluorescence of transverse sections from mouse carotid or thoracodorsal arteries or from a

human skeletal muscle arteriole. In all images, red indicates Hb α expression, green shows internal elastic lamina (IEL) autofluorescence, and blue specifies nuclei. **e**, En face images of Hb α (red) and Hb β (green) expression in ECs from thoracodorsal or carotid arteries. Blue staining represents nuclei. **f**, Western blot analysis of thoracodorsal artery, EC, MEJ, SMC or red blood cell lysates that were chemically crosslinked using BS₃ (bis(sulphosuccinimidyl)suberate) to determine the quaternary structure of Hb α . **g**, mRNA analysis from EC, MEJ and SMC lysates isolated from VCCCs, thoracodorsal and carotid arteries. 18S was used as a normalization factor. In **a–c** and **g** open bars represent *in vitro* data and striped bars indicate *ex vivo* data. Scale bars, 2 μm (**a**), 0.5 μm (**b**), 30 μm (**d**, thoracodorsal artery and carotid) and 10 μm (**d**, human skeletal muscle artery, **e**). NS, not significant. *P* values are shown for each comparison, by analysis of variance (ANOVA) with Bonferroni post-analysis or Student's *t*-test. All error bars represent s.e.m.

and eNOS are in a macromolecular complex and can form a direct protein–protein interaction.

Hb α probably interacts with eNOS to regulate blood vessel tone by controlling NO diffusion through its scavenging by haem iron^{13,17–19}. We studied the mechanism of interaction by measuring loss of NO radical in thoracodorsal and carotid arteries, and in the VCCC model. NO was lost in thoracodorsal arteries, but not carotid arteries; it was also lost in MEJ fractions, but not in EC or SMC lysates (Supplementary Fig. 12a, b). Next, we knocked down endothelial Hb α in isolated arteries (Fig. 2j) or VCCCs (Fig. 2l) using siRNA. Loss of Hb α increased NO diffusion across the vessel wall (Fig. 2k) and in VCCCs (Fig. 2m). Together, these results indicate that endothelial Hb α can regulate arterial tone through its effects on NO diffusion.

Next, we proposed that Hb α haem iron in the oxygenated Fe²⁺ state should control NO diffusion through a fast reaction ($2.4 \times 10^7 \text{ M}^{-1} \text{ s}^{-1}$)²⁰ resulting in dioxygenation^{21,22}, whereas Fe³⁺ state should permit NO diffusion owing to a slower reaction rate ($3.3 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$)²³. We found that Hb α haem iron resides in both states. First, using ultraviolet–visible spectroscopy, we identified a Soret peak ($\sim 420 \text{ nm}$) and Q bands ($\sim 540\text{--}575 \text{ nm}$) in isolated thoracodorsal arteries consistent with oxygen-bound Hb Fe²⁺, whereas there was no peak in carotid arteries (Fig. 3a). Next, we measured the oxidation state of Fe and found approximately 42% existed in the Fe²⁺ and 58% in the Fe³⁺ state (Fig. 3b). These measurements were sensitive to Hb α siRNA

(Fig. 3b). Consistent with this observation, we found that carbon monoxide (CO) ligated Fe²⁺ haem, resulting in increased NO diffusion across isolated vessels (Supplementary Fig. 12c). When MEJ fractions were studied, we found a Soret peak ($\sim 410 \text{ nm}$) characteristic of the Fe³⁺ state (methaemoglobin) (Fig. 3c). Interestingly, pelleted membranes from MEJ fractions were dark brown, consistent with Fe³⁺ oxidation (Supplementary Fig. 13). We found approximately 32% of Fe existed in the Fe²⁺ state and 68% in the Fe³⁺ state (Fig. 3d), results that were also sensitive to Hb α siRNA (Fig. 3d). In addition, we observed an increase in NO diffusion in VCCCs treated with CO (Supplementary Fig. 12d).

Previous work has demonstrated that the NO–haem Fe³⁺ interaction results in reductive nitrosylation, a mechanism known to generate S-nitrosothiols, which we have shown to be crucial for gap-junction regulation at the MEJ^{5,24,25}. Using *N*-acetylcysteine as a bait reactant on the abluminal side (Supplementary Fig. 14a, c), we also found a notable loss of S-nitrosothiol synthesis after Hb α knockdown in thoracodorsal arteries (Supplementary Fig. 14b) and in VCCCs (Supplementary Fig. 14d). Together, these results suggest that the Hb α haem oxidation state regulates both NO diffusion and bioactivation.

Next we determined the mechanism regulating the Hb α oxidation state. In erythrocytes, CYB5R3, a known methaemoglobin reductase, controls the haem iron oxidation state through the reduction of Fe³⁺ (ref. 6). Using immunofluorescence (Fig. 4a, e (*in vivo*, *in vitro*,

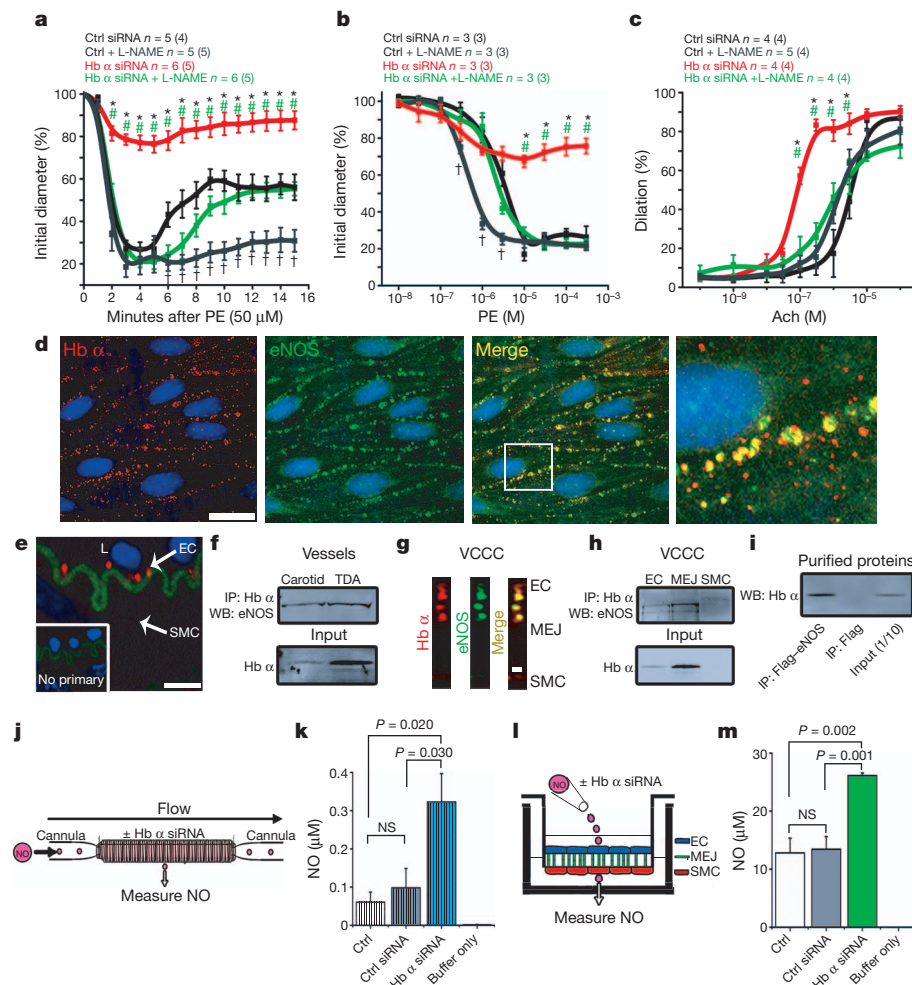
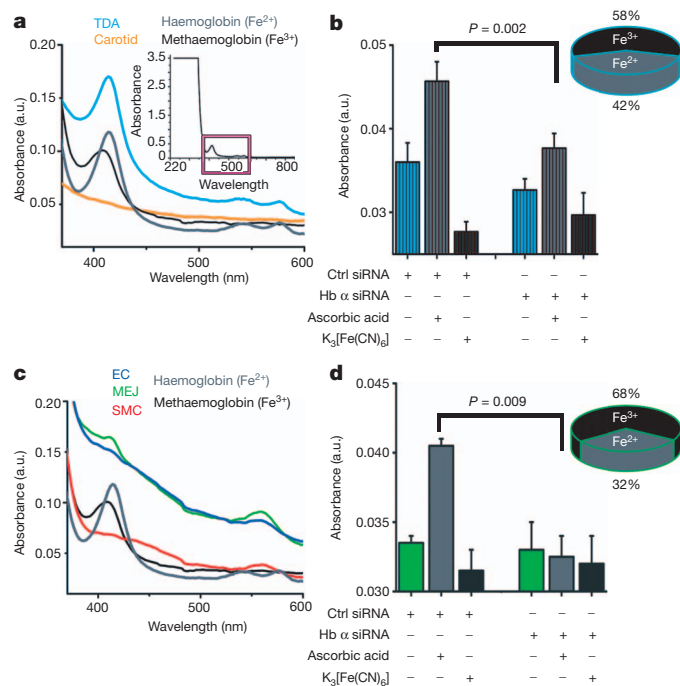


Figure 2 | Hb α regulates vessel tone, NO diffusion and associates with eNOS. **a–c**, Time course effect of 50 μ M phenylephrine (**a**), a dose response to phenylephrine (**b**) and a dose response to ACh (**c**) on thoracodorsal arteries treated with control or Hb α siRNA in the presence or absence of L-NAME. n indicates the number of arteries; value in parenthesis shows number of mice. Asterisk (*) shows significance between control siRNA and Hb α siRNA; hash symbol (#) indicates significance between Hb α siRNA and Hb α siRNA plus L-NAME; dagger (†) represents significance between control plus L-NAME. **d**, *En face* view of a dual immunofluorescence of a mouse thoracodorsal artery showing Hb α (red) and eNOS (green). The white box in the merge panel indicates the region of interest magnified in the far-right panel. **e**, Proximity ligation assay for Hb α and eNOS (red punctates) in transverse mouse thoracodorsal artery sections. Inset shows the negative control. **f**, Western blot (WB) analysis from samples co-immunoprecipitated (IP) for Hb α and probed for eNOS from isolated thoracodorsal and carotid arteries. **g**, Dual immunofluorescence for Hb α and eNOS on transverse section from a VCCC. Red indicates Hb α and green shows eNOS. **h**, Co-immunoprecipitation of Hb α and eNOS expression in VCCC lysates. **i**, Co-immunoprecipitation of purified Flag-eNOS protein probed for Hb α . **j**, Schematic diagram of experimental design illustrating a cannulated vessel transfected with Hb α siRNA showing NO diffusion as a readout. **k**, NO diffusion results from mouse thoracodorsal arteries transfected with control or Hb α siRNA ($n \geq 5$). **l**, Illustration of experimental set up for VCCC experiments. **m**, NO diffusion results from VCCCs transfected with control or Hb α siRNA ($n \geq 4$). Striped bars in **k** represent *ex vivo* data; open bars in **m** indicate *in vitro* data. Scale bars, 10 μ m (**d**, **e**) and 1 μ m (**g**). P values are shown for each comparison, by ANOVA with Bonferroni post-analysis. All error bars represent s.e.m.



respectively), TEM (Fig. 4b), and western blot analysis (Fig. 4c, d (*in vivo*, *in vitro*, respectively)), we identified that CYB5R3 was expressed in ECs and at the MEJ. In addition, we established CYB5R3 is in a complex with Hb α using four separate assays: immunofluorescence (Fig. 4f, g), proximity ligation assay (Fig. 4h), and co-immunoprecipitation from cell lysates and purified proteins (Fig. 4i). Indeed, molecular modelling of the crystal structures for Hb α , eNOS and CYB5R3 revealed a discrete region of high probability where the proteins could interact (Supplementary Fig. 15). Next, we used human CYB5R3 siRNA (knockdown efficiency: ~50%; Supplementary Fig. 16a) and CYB5R3 overexpression to show that CYB5R3 regulates methaemoglobin α reduction. Time-lapse ultraviolet-visible spectrometry demonstrated that the loss of CYB5R3 protein inhibited methaemoglobin α reduction and that overexpression enhanced methaemoglobin α reduction (Supplementary Fig. 16b, c). To determine whether CYB5R3 expression or activity regulates arterial tone, we tested both siRNA directed against endothelial CYB5R3 in thoracodorsal arteries

Figure 3 | The oxidation state of Hb α resides in a mixture of Fe²⁺ and Fe³⁺. **a**, **c**, Ultraviolet-visible spectroscopy analysis of thoracodorsal arteries (**a**) and VCCC fractions (**c**). The inset in **a** indicates the region of interest (magenta box) of the Soret (~420 nm) and Q bands (~540–575 nm). a.u., arbitrary units. **b**, **d**, Measurement of the Hb α oxidation state calculating the ratio of Fe²⁺ to Fe³⁺ in thoracodorsal arteries (**b**; $n = 3$) and VCCC fractions (**d**; $n = 3$) with and without Hb α siRNA. Striped bars in **b** indicate *ex vivo* data; open bars in **d** represent *in vitro* data. K₃[Fe(CN)₆], potassium ferricyanide. P values are indicated for each comparison, by Student's *t*-test. All error bars represent s.e.m.

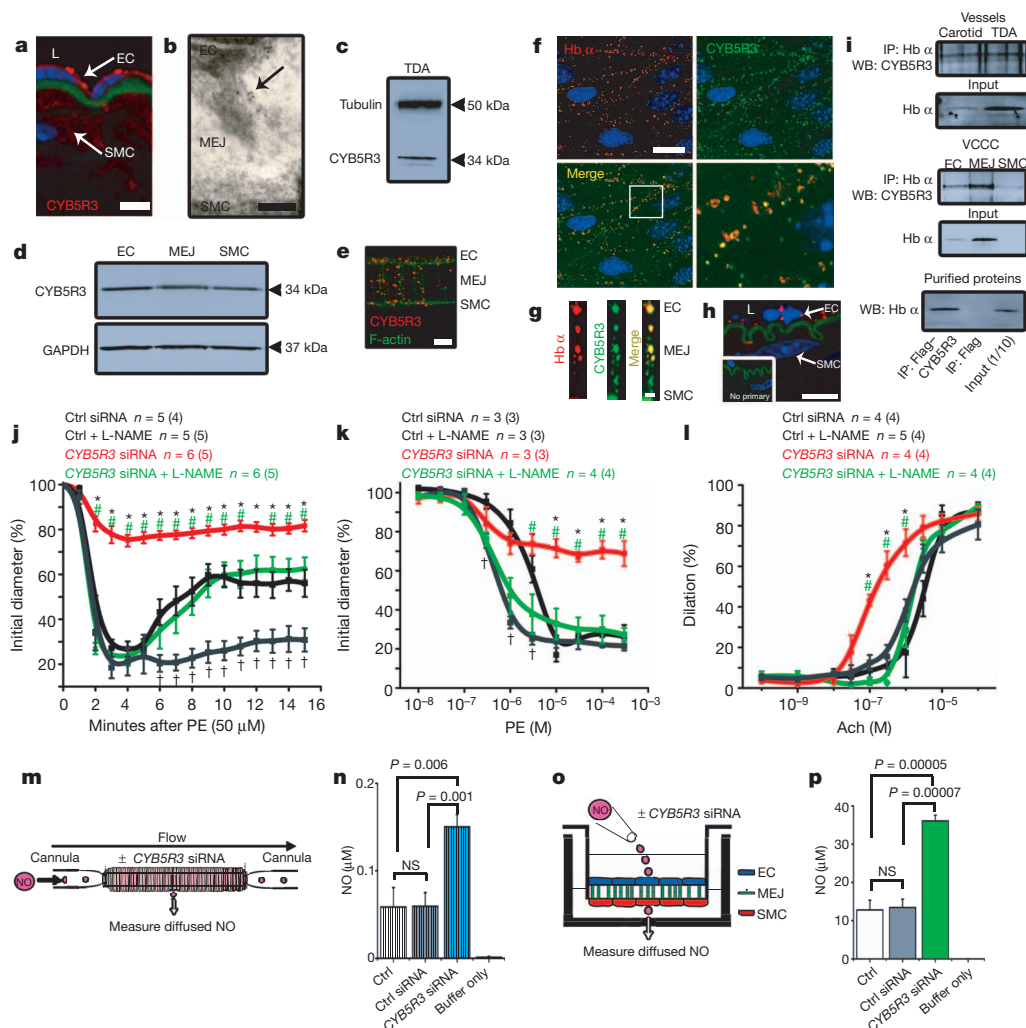


Figure 4 | CYB5R3 expression and activity are crucial for vasomotor tone and NO diffusion. **a**, Immunofluorescence of CYB5R3 expression (red) and nuclei (blue). Green represents autofluorescence from internal elastic lamina. **b**, TEM analysis of CYB5R3 expression at the MEJ (black particles) *in vivo*. **c**, **d**, Western blot analysis of CYB5R3 in thoracodorsal arteries (c) and VCCCs (d). **e**, Immunofluorescence of CYB5R3 expression in VCCCs. Red shows CYB5R3 and green indicates F-actin. **f**, *En face* view of a dual immunofluorescence labelling of a mouse thoracodorsal artery showing Hb α (red) and CYB5R3 (green) in the top panels. White box in the merge image in the bottom left panel shows the region of interest magnified in the bottom right panel. **g**, Colocalization of CYB5R3 (red) and Hb α (green) on a transverse section from the VCCC. **h**, Proximity ligation assay for Hb α and CYB5R3 (red punctates) on transverse mouse thoracodorsal artery sections. Inset shows the negative control. Green shows internal elastic lamina autofluorescence. **i**, Western blot analysis from samples co-immunoprecipitated for Hb α and probed for CYB5R3 from isolated thoracodorsal and carotid arteries, VCCCs or purified proteins. **j-l**, Time course effects of 50 μ M phenylephrine (j), a dose

response to phenylephrine (k) and a dose response to ACh (l) on thoracodorsal arteries treated with control or Hb α siRNA in the presence or absence of L-NAME. *n* indicates the number of arteries; value in parenthesis shows number of mice. Asterisk (*) shows significance between control siRNA and CYB5R3 siRNA; hash symbol (#) indicates significance between CYB5R3 siRNA and CYB5R3 siRNA plus L-NAME; dagger (†) represents significance between control and control plus L-NAME. **m**, Schematic of experimental set up for NO diffusion assay in a cannulated artery that was transfected with CYB5R3 siRNA. **n**, Results from NO diffusion experiment in mouse thoracodorsal arteries with genetic knockdown of CYB5R3 expression ($n \geq 3$). **o**, Illustration showing the experimental set up for VCCC experiments. **p**, NO diffusion results from VCCCs transfected with control or CYB5R3 siRNA ($n = 4$). Striped bars in **n** indicate *ex vivo* data; open bars in **p** represent *in vitro* data. Scale bars, 10 μ m (a, f, h), 0.25 μ m (b), 5 μ m (e) and 1 μ m (g). *P* values are indicated for each comparison, by ANOVA with Bonferroni post-analysis. All error bars represent s.e.m.

and a pharmacological inhibitor of CYB5R3, propylthiouracil (PTU)²⁶. Knockdown efficiency using mouse *Cyb5r3* siRNA was about 70% (Supplementary Fig. 17a). We observed a decrease in arterial reactivity in thoracodorsal arteries transfected with CYB5R3 siRNA after phenylephrine stimulation with a single dose or cumulative concentrations (Fig. 4j, k), and increased arterial reactivity with increasing ACh doses (Fig. 4l). Vascular reactivity to phenylephrine or ACh in thoracodorsal arteries pretreated with PTU is shown in Supplementary Fig. 18a–c. The PTU effect was not reversible with L-thyroxine supplementation after phenylephrine stimulation (Supplementary Fig. 18b, inset). We found no change in the PTU effect after 5-HT addition (Supplementary Table 1). EC₅₀ and E_{max} values are in Supplementary Table 2.

However, after the addition of L-NAME, the effects of CYB5R3 siRNA (Fig. 4j–l) or of PTU treatment (Supplementary Fig. 18a, b) were comparable to control conditions, results that were consistent with Hb α knockdown. We found no difference in basal tone for CYB5R3 siRNA or PTU (Supplementary Fig. 11a, b). Next, we tested the effect of CYB5R3 on NO diffusion in vessels and VCCCs (Fig. 4m, o). Knockdown efficiency with CYB5R3 siRNA was ~30% at the MEJ (Supplementary Fig. 17b) and in the EC monolayer, but not in SMCs (Supplementary Fig. 17c). Both CYB5R3 siRNA and PTU treatment increased NO diffusion across isolated vessels and in VCCCs (Fig. 4n, p and Supplementary Fig. 18d–g). Note that CYB5R3 knockdown did not alter MEJ eNOS or Hb α expression (Supplementary Fig. 17d).

We conclude that EC expression of Hb α has a crucial role in the regulation of NOS-mediated signalling and in the control of arterial vascular reactivity. These results may have far-reaching implications that could influence many aspects of vascular biology and disease. For example, endothelial Hb α expression may participate in blood pressure control, arteriogenesis and anti-inflammatory signalling, as well as affecting other redox signalling molecules (for example, superoxide and hydrogen peroxide). Our results correlate with diagnostic indices for human α -thalassaemia major (*HBA1*^{-/-} *HBA2*^{-/-}) fetuses, who show increased cerebral blood flow during development²⁷. Furthermore, these observations may help to explain why inhibition of CYB5R3 attenuates hypertension²⁸ and may suggest that CYB5R3 is a new therapeutic target for disease treatment. However, studies devoted towards understanding the mechanisms of CYB5R3 regulation and its interaction with Hb α will need to be clarified. More broadly, somatic cell types as diverse as alveolar epithelial cells⁷, macrophages⁹, neurons¹⁰ and renal mesangial cells⁸ express both Hb and NOS. It is thus possible that Hb could regulate NOS signalling pathways relevant to many cell and organ systems. Taken together, these data provide evidence for a new model in which somatic cell Hb oxidation is required for NOS-dependent bioactivity.

METHODS SUMMARY

Human coronary ECs and SMCs were co-cultured and fractionated as previously described²⁹. iTRAQ proteomic screening was used to identify and quantify proteins enriched at the MEJ as previously demonstrated³⁰. Proteins were analysed using western blot, immunofluorescence and immune-TEM. mRNA was measured using real-time-PCR. Isolated thoracodorsal arteries were cannulated, pressurized and stimulated with phenylephrine or ACh as previously shown⁵ or perfused with anaerobic aqueous NO. Detailed methods can be found in Supplementary Methods.

Received 20 January; accepted 27 September 2012.

Published online 31 October 2012.

- Lim, K. H., Ancrile, B. B., Kashatus, D. F. & Counter, C. M. Tumour maintenance is mediated by eNOS. *Nature* **452**, 646–649 (2008).
- Hess, D. T., Matsumoto, A., Kim, S. O., Marshall, H. E. & Stamler, J. S. Protein S-nitrosylation: purview and parameters. *Nature Rev. Mol. Cell Biol.* **6**, 150–166 (2005).
- Bolotina, V. M., Najibi, S., Palacino, J. J., Pagano, P. J. & Cohen, R. A. Nitric oxide directly activates calcium-dependent potassium channels in vascular smooth muscle. *Nature* **368**, 850–853 (1994).
- Shesely, E. G. *et al.* Elevated blood pressures in mice lacking endothelial nitric oxide synthase. *Proc. Natl Acad. Sci. USA* **93**, 13176–13181 (1996).
- Straub, A. C. *et al.* Compartmentalized connexin 43 s-nitrosylation/denitrosylation regulates heterocellular communication in the vessel wall. *Arterioscler. Thromb. Vasc. Biol.* **31**, 399–407 (2011).
- Hultquist, D. E. & Passon, P. G. Catalysis of methaemoglobin reduction by erythrocyte cytochrome B5 and cytochrome B5 reductase. *Nat. New Biol.* **229**, 252–254 (1971).
- Newton, D. A., Rao, K. M., Dluhy, R. A. & Baatz, J. E. Hemoglobin is expressed by alveolar epithelial cells. *J. Biol. Chem.* **281**, 5668–5676 (2006).
- Nishi, H. *et al.* Hemoglobin is expressed by mesangial cells and reduces oxidant stress. *J. Am. Soc. Nephrol.* **19**, 1500–1508 (2008).
- Liu, L., Zeng, M. & Stamler, J. S. Hemoglobin induction in mouse macrophages. *Proc. Natl Acad. Sci. USA* **96**, 6643–6647 (1999).
- Schelshorn, D. W. *et al.* Expression of hemoglobin in rodent neurons. *J. Cereb. Blood Flow Metab.* **29**, 585–595 (2009).
- Halligan, K. E., Jourdain, F. L. & Jourdain, D. Cytochrome is expressed in the vasculature and regulates cell respiration and proliferation via nitric oxide dioxygenation. *J. Biol. Chem.* **284**, 8539–8547 (2009).
- Brunori, M. *et al.* Neuroglobin, nitric oxide, and oxygen: functional pathways and conformational changes. *Proc. Natl Acad. Sci. USA* **102**, 8483–8488 (2005).
- Flogel, U., Merx, M. W., Godecke, A., Decking, U. K. & Schrader, J. Myoglobin: a scavenger of bioactive NO. *Proc. Natl Acad. Sci. USA* **98**, 735–740 (2001).
- Dora, K. A., Doyle, M. P. & Duling, B. R. Elevation of intracellular calcium in smooth muscle causes endothelial cell generation of NO in arterioles. *Proc. Natl Acad. Sci. USA* **94**, 6529–6534 (1997).
- Angelo, M., Hausladen, A., Singel, D. J. & Stamler, J. S. Interactions of NO with hemoglobin: from microbes to man. *Methods Enzymol.* **436**, 131–168 (2008).
- Gladwin, M. T., Lancaster, J. R. Jr, Freeman, B. A. & Schechter, A. N. Nitric oxide's reactions with hemoglobin: a view through the SNO-storm. *Nature Med.* **9**, 496–500 (2003).
- Palmer, R. M., Ferrige, A. G. & Moncada, S. Nitric oxide release accounts for the biological activity of endothelium-derived relaxing factor. *Nature* **327**, 524–526 (1987).
- Ignarro, L. J., Adams, J. B., Horwitz, P. M. & Wood, K. S. Activation of soluble guanylate cyclase by NO-hemoproteins involves NO-heme exchange. Comparison of heme-containing and heme-deficient enzyme forms. *J. Biol. Chem.* **261**, 4997–5002 (1986).
- Ignarro, L. J., Byrns, R. E., Buga, G. M. & Wood, K. S. Endothelium-derived relaxing factor from pulmonary artery and vein possesses pharmacologic and chemical properties identical to those of nitric oxide radical. *Circ. Res.* **61**, 866–879 (1987).
- Cassoly, R. & Gibson, Q. Conformation, co-operativity and ligand binding in human hemoglobin. *J. Mol. Biol.* **91**, 301–313 (1975).
- Doyle, M. P. & Hoekstra, J. W. Oxidation of nitrogen oxides by bound dioxygen in hemoproteins. *J. Inorg. Biochem.* **14**, 351–358 (1981).
- Eich, R. F. *et al.* Mechanism of NO-induced oxidation of myoglobin and hemoglobin. *Biochemistry* **35**, 6976–6983 (1996).
- Sharma, V. S., Traylor, T. G., Gardiner, R. & Mizukami, H. Reaction of nitric oxide with heme proteins and model compounds of hemoglobin. *Biochemistry* **26**, 3837–3843 (1987).
- Tejero, J. *et al.* Low NO concentration-dependence of the reductive nitrosylation reaction of hemoglobin. *J. Biol. Chem.* **287**, 18262–18274 (2012).
- Angelo, M., Singel, D. J. & Stamler, J. S. An S-nitrosothiol (SNO) synthase function of hemoglobin that utilizes nitrite as a substrate. *Proc. Natl Acad. Sci. USA* **103**, 8366–8371 (2006).
- Lee, E. & Kariya, K. Propylthiouracil, a selective inhibitor of NADH-cytochrome b5 reductase. *FEBS Lett.* **209**, 49–51 (1986).
- Lam, Y. H. & Tang, M. H. Middle cerebral artery Doppler study in fetuses with homozygous α -thalassaemia-1 at 12–13 weeks of gestation. *Prenat. Diagn.* **22**, 56–58 (2002).
- Fregly, M. J. & Hood, C. I. Physiologic and anatomic effects of propylthiouracil on normal and hypertensive rats. *Circ. Res.* **7**, 486–496 (1959).
- Heberlein, K. R. *et al.* Plasminogen activator inhibitor-1 regulates myoendothelial junction formation. *Circ. Res.* **106**, 1092–1102 (2010).
- Davalos, A. *et al.* Quantitative proteomics of caveolin-1-regulated proteins: characterization of polymerase I and transcript release factor/CAVIN-1 in endothelial cells. *Mol. Cell Proteomics* **9**, 2109–2124 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the Advanced Microscopy and Histology core at the University of Virginia and the Yale Proteomic Facility. We acknowledge V. Balasubramaniam, S. Lewis and D. Singel for discussions of the data and B. Duling for critical evaluation of experiments and the manuscript. We also thank M. Weiss for the Hb α -stabilizing protein antibody and for discussions. This work was supported by an American Heart Association Scientist Development Grant (B.E.I.), National Institutes of Health grants HL088554 (B.E.I.), HL107963 (B.E.I.), HL059337 (B.G.), HL101871 (B.G.), HL112904 (A.C.S.) and HL007284 (A.W.L. and A.C.S.). M.B. and S.R.J. were supported by American Heart Association postdoctoral fellowships, and A.W.L. and M.Y.L. were supported by American Heart Association predoctoral fellowships.

Author Contributions A.C.S. performed most of the experiments and data analysis. A.W.L. performed vessel transfections. Vascular reactivity was executed by A.W.L. and M.B. S.R.J. carried out immunofluorescence studies and S.T.D. assisted in NO diffusion and consumption assays. M.Y.L. and P.S.B. performed real-time PCR experiments, and A.K.B. helped with all cell culture experiments. L.C. performed the modelling experiments. B.G. helped with experimental design, provided use of the NO analyser and NO, and assisted with final manuscript preparation. B.E.I. initiated, directed and supported the work through all levels of development. All the authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to B.E.I. (brant@virginia.edu).

Structural insight into the type-II mitochondrial NADH dehydrogenases

Yue Feng^{1*}, Wenfei Li^{1*}, Jian Li^{1*}, Jiawei Wang¹, Jingpeng Ge¹, Duo Xu¹, Yanjing Liu², Kaiqi Wu³, Qingyin Zeng², Jia-Wei Wu¹, Changlin Tian^{3,4}, Bing Zhou¹ & Maojun Yang¹

The single-component type-II NADH dehydrogenases (NDH-2s) serve as alternatives to the multisubunit respiratory complex I (type-I NADH dehydrogenase (NDH-1), also called NADH:ubiquinone oxidoreductase; EC 1.6.5.3) in catalysing electron transfer from NADH to ubiquinone in the mitochondrial respiratory chain¹. The yeast NDH-2 (Ndi1) oxidizes NADH on the matrix side and reduces ubiquinone to maintain mitochondrial NADH/NAD⁺ homeostasis. Ndi1 is a potential therapeutic agent for human diseases caused by complex I defects^{2–9}, particularly Parkinson's disease, because its expression restores the mitochondrial activity in animals with complex I deficiency. NDH-2s in pathogenic microorganisms are viable targets for new antibiotics^{10,11}. Here we solve the crystal structures of Ndi1 in its substrate-free, NADH-, ubiquinone- and NADH-ubiquinone-bound states, to help understand the catalytic mechanism of NDH-2s. We find that Ndi1 homodimerization through its carboxy-terminal domain is critical for its catalytic activity and membrane targeting. The structures reveal two ubiquinone-binding sites (UQ_I and UQ_{II}) in Ndi1. NADH and UQ_I can bind to Ndi1 simultaneously to form a substrate–protein complex. We propose that UQ_I interacts with FAD to act as an intermediate for electron transfer, and that NADH transfers electrons through this FAD–UQ_I complex to UQ_{II}. Together our data reveal the regulatory and catalytic mechanisms of Ndi1 and may facilitate the development or targeting of NDH-2s for potential therapeutic applications.

Electron transport chains, which comprise oxidative phosphorylation complexes I to V in the inner membrane of mitochondria, serve as the core components during respiration¹². Complex I catalyses the transfer of two electrons from NADH to reduce ubiquinone to ubiquinol, and is the entry point for a large fraction of the electrons that traverse the respiratory chain^{13,14}. This catalysis is the rate-limiting step in respiration and is central to energy metabolism¹⁵. Three different types of NDH are known: proton-pumping NDH (NDH-1); non-proton-pumping NDH (NDH-2); and the rare sodium-pumping NDH, called NQR^{1,15,16}.

First identified in domestic yeast^{17,18}, NDH-2 is a ~50-kDa, monotopic membrane protein with no energy-transducing capacity. Subsequently, multiple NDH-2s were found in various fungi, plants and primitive eukaryotic cells^{1,19}. Genome-wide analysis has shown that a majority of organisms, such as *Escherichia coli*, contain both NDH-1 and NDH-2. However, some species have only NDH-2 (Supplementary Fig. 1, Supplementary Data 1 and Supplementary Table 1). Ndi1, one of the NDH-2s in *Saccharomyces cerevisiae*, faces the inside surface of the mitochondrial inner membrane, similar to NDH-1²⁰. Here we report the crystal structures of Ndi1 in its apo and substrate-bound states.

Ndi1 forms a globular α/β structure that can be divided into three domains (Fig. 1a and Supplementary Fig. 2). Similar to other NADH- and FAD-associated enzymes, Ndi1 contains two canonical Rossmann

domains with an FAD molecule buried deep in the first one (residues 49–412, referred to as the active domain hereafter) (Fig. 1 and Supplementary Fig. 3). A search of the Dali database²¹ revealed that the NADH- and FAD-associated enzymes are the closest structural homologues of the Ndi1 active domain, with most of them having a root mean squared deviation (r.m.s.d.) of less than 2 Å when compared with the Ndi1 active domain. The structures of Ndi1 and other NADH- and FAD-associated enzymes differ in their C-terminal domains (CTDs; Fig. 1b). Ndi1 CTD is highly conserved among NDH-2s from different species (Supplementary Fig. 2 and Supplementary Data 1 and 2) and contains three β -strands (β 19– β 21) and two α -helices (α 15 and α 16) (Fig. 1a and Supplementary Fig. 2). The two α -helices form a helix–turn–helix motif that extensively interacts with the active domain, whereas the three β -strands make fewer contacts with the active domain (Supplementary Fig. 4).

In all the structures, each asymmetric unit contains an Ndi1 homodimer (Fig. 2a). Interactions between the two Ndi1 monomers are mainly mediated through helix α 16 of the CTD (Fig. 2b and Supplementary Fig. 5a, b), burying a surface area of 5,982 Å². Consistent with this, we eluted wild-type Ndi1 from a gel filtration column with a native molecular mass of ~150 kD, corresponding to the calculated size of an Ndi1 dimer (Supplementary Fig. 6a). The linear relationship between Ndi1 concentration and its activity indicated that a single enzyme catalysed the electron transfer reaction (Supplementary Fig. 6b). To investigate whether Ndi1 was oligomeric *in vivo*, we co-expressed Ndi1–FLAG and Ndi1–Myc proteins in yeast cells and performed immunoprecipitation assays. As shown in Fig. 2c, the two Ndi1 proteins interacted with each other, indicating that Ndi1 formed an oligomer in yeast cells. Deletion of α 16 compromised the interaction between Ndi1–FLAG and Ndi1–Myc, and deletion of both α 15 and α 16 abolished the interaction (Fig. 2c). This result indicated that the CTD is important for Ndi1 oligomerization. We further confirmed the functional importance of the CTD through a rescue assay. The plasmid carrying wild-type Ndi1 but not the mutants with α 16 or α 15 and α 16 deleted, rescued the yeast *ndi1Δ* phenotype (Supplementary Fig. 5c). Thus, Ndi1 functions as a homodimer both *in vitro* and *in vivo*.

The membrane anchoring of *E. coli* NDH-2 is required for its activities²². One side of the Ndi1 homodimer contains a large hydrophobic surface, which is formed by the solvent-exposed residues from the CTD and surrounded by numerous positively charged patches (Fig. 2a, d). In solution, the hydrophobic surface was shielded by the detergent molecules used for protein purification, because several detergent molecules bound to this region in the structures of Ndi1 (Supplementary Fig. 7). This suggested that the hydrophobic surface could have a role in the mitochondrial membrane localization of Ndi1. Indeed, fractionation indicated that wild-type Ndi1 protein was exclusively found in the membrane fraction. In contrast, a large pool of the

¹State Key Laboratory of Biomembrane and Membrane Biotechnology, Tsinghua-Peking Center for Life Sciences, School of Life Sciences, Tsinghua University, Beijing 100084, China. ²State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China. ³Hefei National Laboratory for Physical Sciences at Microscale, School of Life Sciences, University of Science and Technology of China, Hefei, Anhui 230026, China. ⁴High Magnetic Field Laboratory, Chinese Academy of Sciences, Hefei, Anhui 230031, China.

*These authors contributed equally to this work.

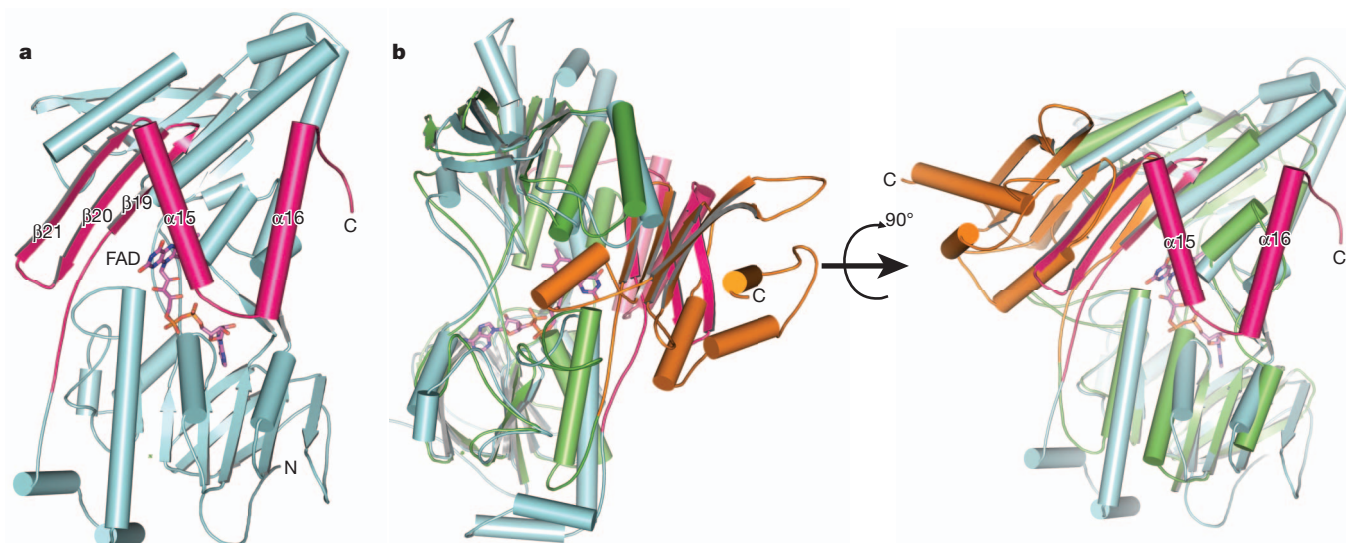


Figure 1 | Ndi1 has a unique CTD. **a**, Overall structure of Ndi1. The CTD (residues 439–513) of Ndi1 is coloured in pink and the active domain in cyan. FAD is shown in stick representation. **b**, Two views of the structural alignment

between Ndi1 and the Pdr–Pdx complex from *Pseudomonas putida* (Protein Data Bank ID, 3LB8). The CTD of Pdr is coloured in orange and the remaining part in green. Colour codes for Ndi1 are the same as those in **a**.

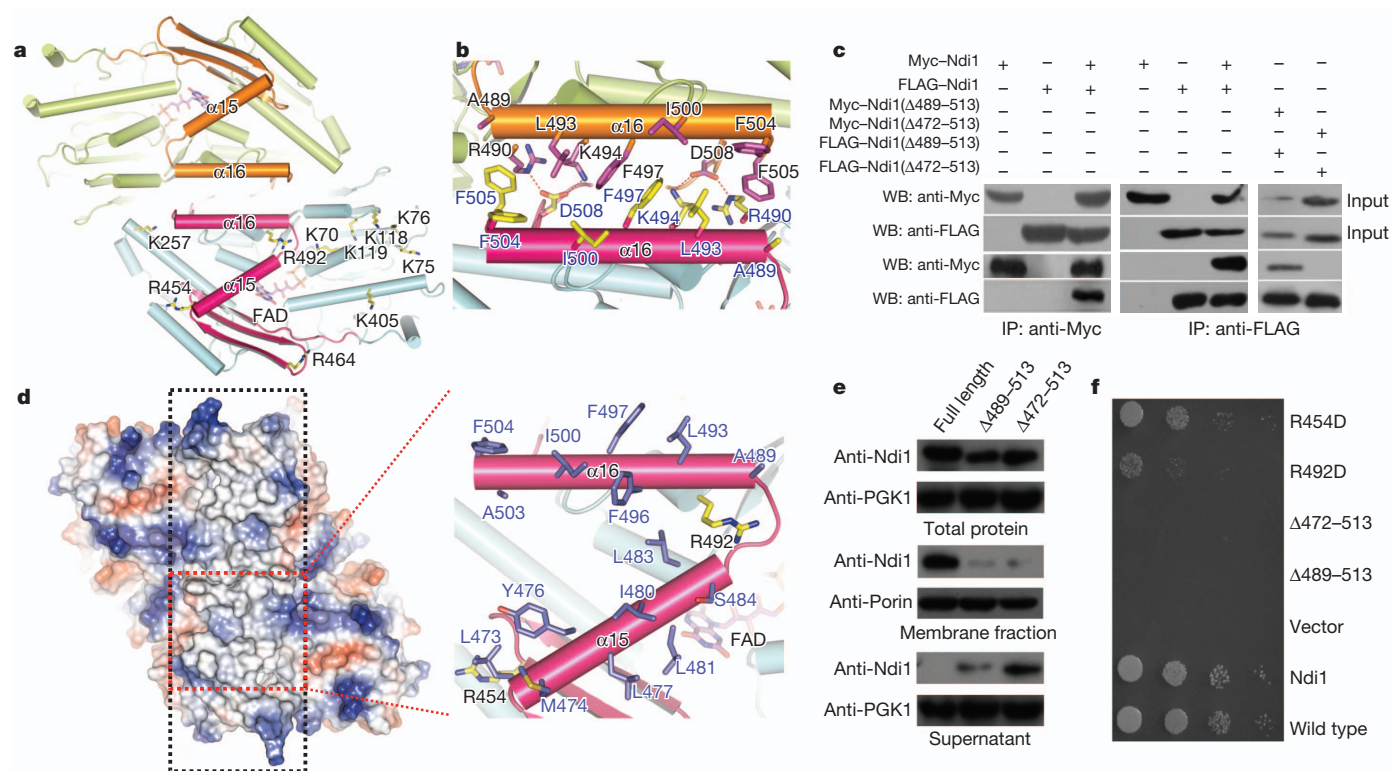


Figure 2 | The CTD mediates Ndi1 homodimerization and membrane attachment. **a**, Ndi1 forms a homodimer in crystals. One protomer is coloured as in Fig. 1a and the other one is coloured in orange (CTD) and lime (remainder). The conserved, positively charged residues on the membrane attachment side of one protomer are shown as yellow sticks. **b**, Detailed interactions between $\alpha 16$ of the two Ndi1 protomers. **c**, Ndi1 forms a dimer *in vivo*. Top: western blots of lysates (input) of *ndi1Δ trp1Δ* cells expressing the wild-type and the mutant forms of Ndi1–Myc, Ndi1–FLAG or both. Proteins were detected by western blot with anti-Myc and anti-FLAG antibodies. Bottom: Cell lysates in the top panels were subjected to anti-FLAG or anti-Myc immunoprecipitation (IP) followed by anti-Myc and anti-FLAG immunoblotting (WB). **d**, A large hydrophobic surface formed by the CTD in

the Ndi1 homodimer. Left: electrostatic surface of the Ndi1 homodimer. The hydrophobic surface is highlighted with a dashed box. Right: the side chains of the surface residues from $\alpha 15$ and $\alpha 16$ of the CTD. For clarity, only one Ndi1 monomer is shown. **e**, The CTD is essential for Ndi1 membrane attachment. The lysates of *ndi1Δ* cells transformed with vectors encoding the indicated forms were fractionated into three samples: total protein, membrane fraction and supernatant. The samples were subjected to western blots. The loading amounts of the samples were examined using anti-Porin for the membrane fractions and anti-PGK1 antibody for the total protein and supernatant fractions. **f**, Mutations compromising Ndi1 membrane attachment cause growth defects in yeast cells.

mutant Ndi1 protein with $\alpha 15$ and $\alpha 16$ deleted was detected in the supernatant fraction (Fig. 2e), supporting a role for the two α -helices in Ndi1 membrane anchoring. The two α -helices were required for yeast viability, because their deletion resulted in a lethal phenotype (Fig. 2f).

We then solved the structure of the NADH-bound Ndi1. As observed in other NADH and FAD reductases, NADH binds to the second Rossmann domain and is fully accessible to FAD from the matrix side (Supplementary Fig. 8). Stabilized by a number of hydrogen-bonding interactions, NADH binding causes a slight conformational change in this domain (Supplementary Fig. 9a, b). Supporting their functional importance, single mutations of the NADH-interacting residues caused a lethal phenotype in yeast (Supplementary Fig. 9c). The nicotinamide ring of NADH is orientated in parallel with the *re*-face of the FAD isoalloxazine ring, with the distance between C4 of NADH and N5 of FAD being 3 Å, indicating that a direct electron transfer between them can occur.

The structure of the ubiquinone-bound form was determined by co-crystallizing Ndi1 with ubiquinone in the absence of NADH. One ubiquinone molecule (UQ_I) with a well-defined density is located close to FAD (Fig. 3a) and lies in the groove formed by $\beta 19$ – $\beta 21$ and $\alpha 15$ from the CTD (Supplementary Figs 10a, b). The distance between the centre of the FAD isoalloxazine ring and UQ_I is about 7.8 Å. This site was occupied by a Triton X-100 (TRT) molecule in the apo structure (Supplementary Fig. 7), explaining why the protein extracted with *n*-dodecyl- β -D-maltopyranoside (DDM) but not with TRT contained bound ubiquinone^{23,24}. Mutations of the residues that interact with UQ_I markedly attenuated the enzymatic activity of Ndi1 and

compromised yeast survival (Supplementary Figs 10c, d). Although comparatively less well defined, a second ubiquinone molecule (UQ_{II}) was assigned to the stretch of density adjacent to UQ_I (Fig. 3a). This assignment was consistent with the recent studies indicating that Ndi1 contains both a bound ubiquinone and a catalytic ubiquinone at similar sites^{24,25}. Both ubiquinone molecules are located on the *si*-face of the isoalloxazine ring of FAD and extend their aliphatic side chains towards the large hydrophobic surface (Supplementary Fig. 10a). The less well-defined UQ_{II} forms fewer contacts with Ndi1 than does UQ_I (Supplementary Fig. 10b).

Putidaredoxin reductase (Pdr) is an NADH-dependent enzyme and transfers electrons to P450 through forming a complex with the [2Fe–2S]-containing protein putidaredoxin²⁶ (Pdx). As anticipated, the structure of Pdr resembles that of the Ndi1 active domain. UQ_{II} in Ndi1 is similarly positioned relative to the [2Fe–2S] cluster in Pdx. The distance between the centre of the isoalloxazine ring and the UQ_{II} centre is about 11.2 Å, comparable to the 12 Å distance between FAD and the [2Fe–2S] cluster in the Pdr–Pdx complex (Supplementary Fig. 11). These structural observations suggest that it is spatially possible for the electron to be transferred from reduced FAD ($FADH_2$) to UQ_{II} .

Previous studies suggested that Ndi1 could form a ternary complex with NADH and ubiquinone²³. We therefore soaked the ubiquinone-bound Ndi1 crystals with NADH and solved the structure of the Ndi1–NADH–ubiquinone complex (Fig. 3b). UQ_{II} was not observed in the complex. Because the crystals of the complex were obtained in the presence of oxygen, it might represent an enzyme–substrate complex

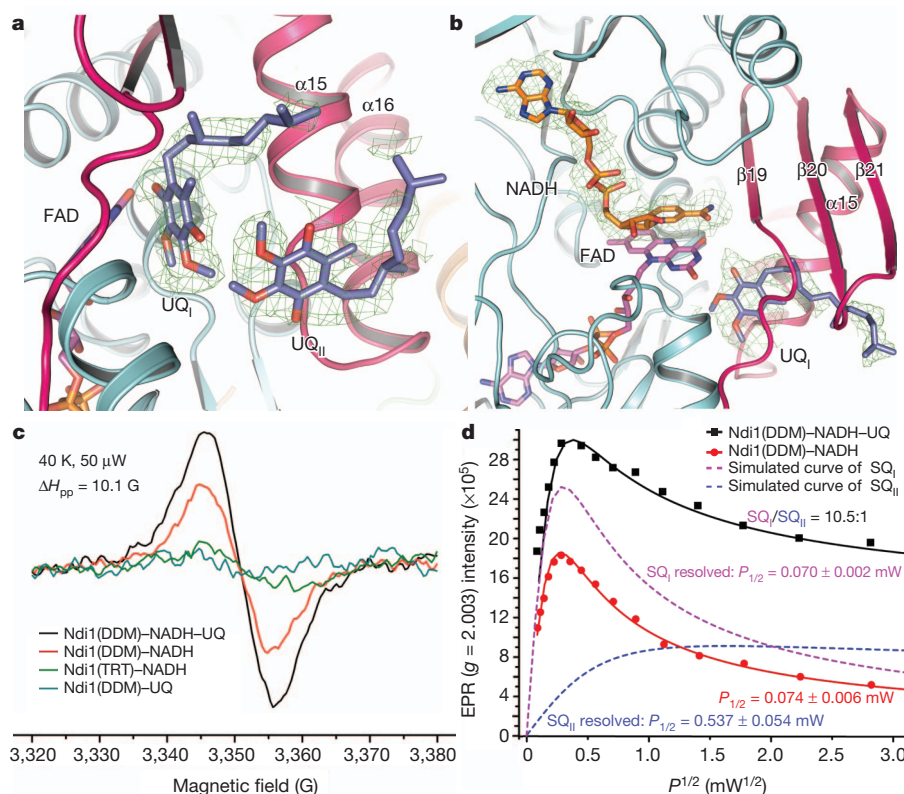


Figure 3 | Electron transfer of Ndi1 involves two ubiquinone molecules. **a**, Two ubiquinone molecules bind to Ndi1. The ubiquinone-binding sites in the Ndi1–ubiquinone complex are shown in ribbon representation. The ubiquinone and FAD molecules are shown as sticks and labelled. The $F_o - F_c$ omit density (green) is contoured at 3.5 σ . **b**, Ribbon representation of the reaction centre in the Ndi1–NADH–ubiquinone complex. The $F_o - F_c$ omit density (green) is contoured at 3.5 σ . **c**, Generation of Ndi1-catalysed radicals requires both NADH and ubiquinone. The EPR spectra of the $g = 2.003$ signals (g , electron g -factor) arise from ubisemiquinones of Ndi1 proteins as indicated.

ΔH_{pp} , peak-to-peak width. **d**, Ndi1-catalysed generation of two populations of radicals. The power saturation behaviours of the $g = 2.003$ EPR signals of Ndi1 were monitored at 40 K. All experimental parameters except the microwave power were the same as in **c**. The ubisemiquinone signal of Ndi1(DDM)–NADH (red) was fitted by the equation described in Methods, and that of the sample with excessive ubiquinone (black) was resolved as the sum of two components. The theoretical curves for these two components are shown by the purple and blue dotted lines. Two sets of half-saturation parameters ($P_{1/2}$) and the relative concentration ratio are indicated.

rather than a metastable enzyme–intermediate complex. Comparison with the apo-Ndi1 structure shows that Ndi1 undergoes almost no conformational changes when bound by ubiquinones (r.m.s.d, 0.22 Å) or NADH–ubiquinone (r.m.s.d, 0.26 Å). The UQ_I -binding site is almost identical in the two structures, suggesting that UQ_I may act as a cofactor together with FAD to mediate electron transfer (Supplementary Fig. 12).

Our structures reveal that Ndi1 possesses a unique CTD (Fig. 1 and Supplementary Fig. 2) that mediates the Ndi1 dimerization required for its activity (Figs 2a–c and Supplementary Fig. 5). The CTD-mediated Ndi1 dimerization creates a large hydrophobic surface at one side (Fig. 2d) and a hydrophilic groove at the other (Supplementary Fig. 13). The biochemical assay (Figs 2e, f) supports there being a role for the hydrophobic surface in targeting Ndi1 to the mitochondrial membrane. The positively charged patches around the hydrophobic surface could also contribute to the membrane localization of Ndi1, presumably through interactions with the phosphate groups of membrane-localized phospholipids. Further investigations are needed to understand the specific mitochondrial membrane targeting of Ndi1. Additionally, the CTD is directly involved in the recognition of ubiquinones by Ndi1 (Figs 3a, b). Taken together, we conclude that the unique CTD of the single-component enzyme Ndi1 enables it to perform the electron transfer function of the multi-component NDH-1s.

Like many ubiquinone-reactive enzymes^{27–30}, Ndi1 has two ubiquinone-binding sites (Fig. 3a). UQ_I may not readily exchange with the ubiquinone pool, because it forms extensive interactions with the CTD of Ndi1. In contrast, UQ_{II} may better serve this purpose because it has much fewer contacts with Ndi1 (Supplementary Figs 10a, b). Our structures suggest that there is an electron transfer pathway from NADH to UQ_{II} , with UQ_I acting as an intermediate together with FAD. FAD first accepts two electrons from NADH to form $FADH_2$. Several possible mechanisms could be responsible for subsequent electron transfer. $FADH_2$ could transfer one electron to UQ_I and one to UQ_{II} , reducing both to ubisemiquinones. Then a disproportionation reaction between them could occur, restoring UQ_I to its previously oxidized state and converting UQ_{II} into ubiquinol. Alternatively, $FADH_2$ could transfer the electrons one by one to UQ_{II} through the bound UQ_I , reducing UQ_{II} to ubiquinol. It is also possible that $FADH_2$ transfers both electrons successively to UQ_I and that the two electrons are then further transferred to UQ_{II} consecutively. The net results of these three different pathways are the same: two electrons from NADH are transferred through the FAD– UQ_I complex to UQ_{II} . All three mechanisms involve ubisemiquinone radicals. Indeed, we detected the NADH–ubiquinone-dependent ubisemiquinone radicals in the Ndi1-catalysed electron transfer reaction using electron paramagnetic resonance (EPR) (Fig. 3c). Moreover, the power saturation experiments suggested the existence of two different populations of ubisemiquinone radicals (Fig. 3d). Future studies, such as EPR experiments at different pHs and redox titrations, are needed to establish the detailed electron transfer pathway of Ndi1.

Our study demonstrates that the electron transfer activity of the single-component Ndi1 is enabled by its unique CTD. Our data also support a model in which Ndi1 catalyses the electron transfer from NADH to UQ_{II} through a mechanism mediated by the FAD– UQ_I intermediate. This information could be valuable in guiding the engineering of Ndi1 for therapy and the design of antiparasitic inhibitors of NDH-2s.

METHODS SUMMARY

The wild-type and mutant forms of Ndi1 were expressed and purified from *E. coli* strain C43. Crystals were grown at 18 °C by the hanging-drop vapour diffusion method. Diffraction data sets were collected at beamline BL17U of the Shanghai Synchrotron Radiation Facility. The structures were solved by single-wavelength anomalous diffraction and molecular replacement methods. Data collection and structure refinement statistics are summarized in Supplementary Table 2. The

EPR data were collected at the High Magnetic Field Laboratory of Chinese Academy of Sciences.

Full Methods and any associated references are available in the online version of the paper.

Received 29 April; accepted 29 August 2012.

Published online 21 October 2012.

- Melo, A. M., Bandejas, T. M. & Teixeira, M. New insights into type II NAD(P)H:quinone oxidoreductases. *Microbiol. Mol. Biol. Rev.* **68**, 603–616 (2004).
- Sanz, A. *et al.* Expression of the yeast NADH dehydrogenase Ndi1 in *Drosophila* confers increased lifespan independently of dietary restriction. *Proc. Natl Acad. Sci. USA* **107**, 9105–9110 (2010).
- Maas, M. F., Sellem, C. H., Krause, F., Dencher, N. A. & Sainsard-Chanet, A. Molecular gene therapy: overexpression of the alternative NADH dehydrogenase NDI1 restores overall physiology in a fungal model of respiratory complex I deficiency. *J. Mol. Biol.* **399**, 31–40 (2010).
- Perales-Clemente, E. *et al.* Restoration of electron transport without proton pumping in mammalian mitochondria. *Proc. Natl Acad. Sci. USA* **105**, 18735–18739 (2008).
- Marella, M., Seo, B. B., Flotte, T. R., Matsuno-Yagi, A. & Yagi, T. No immune responses by the expression of the yeast Ndi1 protein in rats. *PLoS ONE* **6**, e25910 (2011).
- Marella, M., Seo, B. B., Yagi, T. & Matsuno-Yagi, A. Parkinson's disease and mitochondrial complex I: a perspective on the Ndi1 therapy. *J. Bioenerg. Biomembr.* **41**, 493–497 (2009).
- Marella, M., Seo, B. B., Thomas, B. B., Matsuno-Yagi, A. & Yagi, T. Successful amelioration of mitochondrial optic neuropathy using the yeast NDI1 gene in a rat animal model. *PLoS ONE* **5**, e11472 (2010).
- Barber-Singh, J., Seo, B. B., Matsuno-Yagi, A. & Yagi, T. Protective role of rAAV-NDI1, serotype 5, in an acute MPTP mouse Parkinson's model. *Parkinson's Dis.* **2011**, 438370 (2011).
- Yagi, T. *et al.* Can a single subunit yeast NADH dehydrogenase (Ndi1) remedy diseases caused by respiratory complex I defects? *Rejuvenation Res.* **9**, 191–197 (2006).
- Teh, J. S., Yano, T. & Rubin, H. Type II NADH: menaquinone oxidoreductase of *Mycobacterium tuberculosis*. *Infect. Disord. Drug Targets* **7**, 169–181 (2007).
- Biagini, G. A. *et al.* Generation of quinolone antimalarials targeting the Plasmodium falciparum mitochondrial respiratory chain for the treatment and prophylaxis of malaria. *Proc. Natl Acad. Sci. USA* **109**, 8298–8303 (2012).
- Rich, P. R. & Marechal, A. The mitochondrial respiratory chain. *Essays Biochem.* **47**, 1–23 (2010).
- Efremov, R. G. & Sazanov, L. A. Structure of the membrane domain of respiratory complex I. *Nature* **476**, 414–420 (2011).
- Walker, J. E. The NADH:ubiquinone oxidoreductase (complex I) of respiratory chains. *Q. Rev. Biophys.* **25**, 253–324 (1992).
- Kerscher, S., Drose, S., Zickermann, V. & Brandt, U. The three families of respiratory NADH dehydrogenases. *Results Probl. Cell Differ.* **45**, 185–222 (2008).
- Barquera, B., Zhou, W., Morgan, J. E. & Gennis, R. B. Riboflavin is a component of the Na⁺-pumping NADH-quinone oxidoreductase from *Vibrio cholerae*. *Proc. Natl Acad. Sci. USA* **99**, 10322–10324 (2002).
- Ohnishi, T., Kawaguchi, K. & Hagihara, B. Preparation and some properties of yeast mitochondria. *J. Biol. Chem.* **241**, 1797–1806 (1966).
- Ohnishi, T., Sottocasa, G. & Ernster, L. Current approaches to the mechanism of energy-coupling in the respiratory chain. Studies with yeast mitochondria. *Bull. Soc. Chim. Biol. (Paris)* **48**, 1189–1203 (1966).
- Kerscher, S. J. Diversity and origin of alternative NADH:ubiquinone oxidoreductases. *Biochim. Biophys. Acta* **1459**, 274–283 (2000).
- Seo, B. B. *et al.* Molecular remedy of complex I defects: rotenone-insensitive internal NADH-quinone oxidoreductase of *Saccharomyces cerevisiae* mitochondria restores the NADH oxidase activity of complex I-deficient mammalian cells. *Proc. Natl Acad. Sci. USA* **95**, 9167–9171 (1998).
- Holm, L. & Rosenstrom, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545–W549 (2010).
- Villegas, J. M., Volentini, S. I., Rintoul, M. R. & Rapisarda, V. A. Amphipathic C-terminal region of *Escherichia coli* NADH dehydrogenase-2 mediates membrane localization. *Arch. Biochem. Biophys.* **505**, 155–159 (2011).
- Yang, Y. *et al.* Reaction mechanism of single subunit NADH-ubiquinone oxidoreductase (Ndi1) from *Saccharomyces cerevisiae*: evidence for a ternary complex mechanism. *J. Biol. Chem.* **286**, 9287–9297 (2011).
- Yamashita, T., Nakamaru-Ogiso, E., Miyoshi, H., Matsuno-Yagi, A. & Yagi, T. Roles of bound quinone in the single subunit NADH-quinone oxidoreductase (Ndi1) from *Saccharomyces cerevisiae*. *J. Biol. Chem.* **282**, 6012–6020 (2007).
- Murai, M. *et al.* Characterization of the ubiquinone binding site in the alternative NADH-quinone oxidoreductase of *Saccharomyces cerevisiae* by photoaffinity labeling. *Biochemistry* **49**, 2973–2980 (2010).
- Sevrioukova, I. F., Poulos, T. L. & Churbanova, I. Y. Crystal structure of the putidaredoxin reductase x putidaredoxin electron transfer complex. *J. Biol. Chem.* **285**, 13616–13620 (2010).
- Ingledew, W. J., Ohnishi, T. & Salerno, J. C. Studies on a stabilisation of ubisemiquinone by *Escherichia coli* quinol oxidase, cytochrome *bo*. *Eur. J. Biochem.* **227**, 903–908 (1995).
- Adelroth, P., Paddock, M. L., Sagie, L. B., Feher, G. & Okamura, M. Y. Identification of the proton pathway in bacterial reaction centers: both protons associated with

- reduction of QB to QBH₂ share a common entry point. *Proc. Natl Acad. Sci. USA* **97**, 13086–13091 (2000).
29. Okamura, M. Y., Paddock, M. L., Graige, M. S. & Feher, G. Proton and electron transfer in bacterial reaction centers. *Biochim. Biophys. Acta* **1458**, 148–163 (2000).
 30. Ohnishi, T., Ohnishi, S. T., Shinzawa-Itoh, K., Yoshikawa, S. & Weber, R. T. EPR detection of two protein-associated ubiquinone components (SQ(Nf) and SQ(Ns)) in the membrane in situ and in proteoliposomes of isolated bovine heart complex I. *Biochim. Biophys. Acta* **1817**, 1803–1809 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We would like to thank the staff at beamline BL17U of the Shanghai Synchrotron Radiation Facility for their assistance with data collection. We thank W. Tong for maintenance and support of the EPR facility at the High Magnetic Field Laboratory, Chinese Academy of Sciences. This work was supported by the Ministry of Science and Technology of China (2011CB910502, 2011CB910900 and

2012CB911101), the National Natural Science Foundation of China (31030020 and 31170679) and Chinese Key Research Plan-Protein Sciences (2011CB911104).

Author Contributions M.Y. designed and directed the project. Y.F., W.L., J. W., J.G., D.X. and J.-W.W. purified the proteins, grew the crystals, collected data, solved the crystal structures and performed the *in vitro* activity analyses. Y.L. and Q.Z. performed the genome analysis of the NDHs. B.Z. and J.L. performed the *in vivo* biological analyses. Y.F., K.W. and C.T. performed the EPR analyses. M.Y. analysed the data and wrote the paper with the help of all the authors.

Author Information The atomic coordinates and structure factors of the apo-, NADH-, ubiquinone- and NADH-ubiquinone-bound forms of Ndi1 have been deposited in the Protein Data Bank under accession codes 4G6G, 4G6H, 4G74 and 4G73, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.Y. (maojunyang@tsinghua.edu.cn).

METHODS

Genome-wide analysis of phylogenetic relationships. NDH sequences from representative species of Prokaryota, Protista, Plantae, Fungi and Metazoa were used to investigate the evolutionary history of this gene family. Because some of the selected species were available only with draft genomes, two methods were used to obtain the sequences. The genes from completely sequenced genomes were obtained from KEGG (the Kyoto Encyclopedia of Genes and Genomes) ORTHOLOGY Database (KEGG ORTHOLOGY Entry K03885). (The locus tags of NDHs are listed in Supplementary Table 1.) The genes from draft genomes (*Saccharomyces mikatae* and *Saccharomyces bayanus*; see Supplementary Data 2) were identified by TBLASTN searches using three full-length NDH protein sequences of *S. cerevisiae*. The collected sequences were aligned using the MUSCLE (multiple sequence comparison by log-expectation) program³¹. The alignment was further adjusted manually using BIOEDIT³². Phylogenetic relationships among the NDHs were reconstructed using a maximum-likelihood procedure in PHYML³³ with the Jones–Taylor–Thornton amino-acid substitution model and 100 bootstrap replicates. (The tree and alignment are available as Supplementary Fig. 1 and Supplementary Data 1 and 2.) Furthermore, we obtained the subunit numbers of NADH:ubiquinone oxidoreductase (complex I) of each species from the KEGG PATHWAY Database (pathway entry, ko00190).

Protein expression and purification. The complementary DNA of Ndi1 from *S. cerevisiae* was cloned into pQE-80L vector. The Ndi1 mutants were generated by two-step PCR and were subcloned, overexpressed and purified in the same way as wild-type protein. Selenomethionine (Se-Met)-labelled protein was purified similarly. Overexpression of Ndi1 was induced in *E. coli* strain C43 using 0.5 mM isopropyl- β -D-thiogalactopyranoside when the cell density reached $D_{600\text{ nm}} = 1.0$. After growth for 10–12 h at 35 °C and 80 r.p.m., the cells were collected, homogenized in a buffer containing 50 mM Tris (pH 8.0) and 1 mM EDTA, and disrupted by sonication. Cell debris was removed by low-speed centrifugation for 15 min. The supernatant was collected and applied to ultracentrifugation at 150,000g for 1 h. The membrane fraction was collected and suspended in 50 mM Tris (pH 7.6), 200 mM NaCl, 0.1 mM EDTA and 10% glycerol (buffer A). After the addition of Triton X-100 at a final concentration of 0.3% (w/v) and 1 mM PMSF, the sample was incubated for 2 h with slow stirring at 4 °C. Following another ultracentrifugation step at 150,000g for 30 min, the supernatant was collected and loaded onto Ni²⁺-nitrilotriacetate affinity resin (Ni-NTA, Qiagen) pre-equilibrated with buffer A containing 0.1% Triton X-100. Subsequently, the resin was washed three times, each time with 5 ml of buffer A containing 0.1% Triton X-100 and 8 mM histidine. Then the protein was eluted with buffer A containing 0.02% Triton X-100 and 200 mM histidine. The purified Ndi1 protein was concentrated and subjected to gel filtration chromatography (Superdex-200 10/30, GE Healthcare) in a buffer containing 25 mM Tris (pH 7.6), 300 mM NaCl, 5 mM DTT, 0.1 mM EDTA and 0.02% Triton X-100. The peak fractions were collected for biochemical analyses or crystallization.

Crystallization, data collection and structure determination. Crystals were grown at 18 °C by the hanging-drop method by mixing 1 μ l of protein (10 mg ml⁻¹) with 1 μ l of reservoir solution (3.8 M potassium formate, 2% w/v polyethylene glycol monomethyl ether 2,000 (pH 4.2 or 10.2)). The crystals appeared overnight and grew to full size in about four to five days. To obtain the NADH-bound Ndi1 crystals, we soaked the crystals in cryo-protectant with 10 mM NADH for about 1–3 min. To obtain the ubiquinone-bound Ndi1 crystals, we made fresh 50 mM ubiquinone-4 (UQ₄) stock solution in DMSO. Ndi1 and ubiquinone were co-crystallized with a final concentration of 0.5 mM UQ₄. To get the NADH-ubiquinone-bound Ndi1 crystals, we soaked the ubiquinone-bound Ndi1 crystals in cryo-protectant with 10 mM NADH for about 1–3 min. The crystals were then transferred to liquid nitrogen.

All the data were collected at Shanghai Synchrotron Radiation Facility beamline BL17U and were integrated and scaled using the HKL2000 package³⁴. Further processing was carried out using programs from the CCP4 suite³⁵. Data collection statistics are summarized in Supplementary Table 2. The self-rotation function with POLARREN revealed a prominent peak at the section of $k = 180^\circ$, indicating the presence of a non-crystallographic two-fold symmetry axis. SHELXD³⁶ was used to locate the positions of selenium sites, and HA_NCS³⁷ independently identified a two-fold non-crystallographic symmetry axis from the anomalous sites. The non-crystallographic symmetry from selenium sites has the same orientation as that from the self-rotation function, which verified the validity of the anomalous scatter positions. The identified anomalous scattering sites were input to PHASER³⁸ for single-wavelength anomalous diffraction phasing. The real-space constraints were applied to the electron density map in DM³⁹. The final electron density map was of sufficient quality for BUCCANEER⁴⁰ to be able to build almost the complete model. The final model rebuilding was performed using COOT⁴¹ and the protein structure was refined with PHENIX⁴² against the high-resolution native data using non-crystallographic symmetry and stereochemistry information as

restraints. Other structures were solved using molecular replacement. Data collection and structure refinement statistics are summarized in Supplementary Table 2.

Yeast strains and plasmids. The strains used in this study were derived from haploid strain BY4742 (*MAT α his3 Δ 1 leu2 Δ 0 lys2 Δ 0 ura3 Δ 0*). *ndi1 Δ* was generated by disruption of the *NDI1* gene through homologous recombination with KanMX4 as the replacement marker. Screen of positive colonies was achieved by spreading cells on YPD (yeast extract/peptone/dextrose) medium plus 0.4 mg ml⁻¹ G418. Growth defects on YPG (yeast extract/peptone/glycerol) medium were used to confirm the disruption. *ndi1 Δ trp1 Δ* was generated by replacing the *TRP1* gene of *ndi1 Δ* with *LEU2*. After selection on the SD-Leu medium, the colonies were further streaked on SD-Trp plates to confirm the deletion.

The vector pADH1–YES2 was modified from the pYES2 vector from Invitrogen, in which the *GAL1* promoter was replaced with an *ADH1* promoter for the purpose of constitutive expression of the target gene. pYES2–Ndi1 and pADH1–YES2–Ndi1 were generated by insertion of the complementary DNA of Ndi1 between the restriction sites EcoRI and XhoI of the vectors pYES2 and pADH1–YES2, respectively. The Ndi1 mutants were generated by two-step PCR and were subcloned into the two vectors using the same approach. To obtain pYES2–Myc–Ndi1 and pYES3–FLAG–Ndi1 constructs, DNA sequences corresponding to Myc (EQKLISEEDL) and FLAG (DYKDDDDK) tags were inserted into the complementary DNA sequence of Ndi1 of pYES2–Ndi1 and pYES3–Ndi1 constructs, respectively, just after its mitochondrial targeting sequence.

Yeast growth assay. For growth on agar plates, yeasts expressing wild-type and mutated *NDI1* genes in pYES2 or pADH1–YES2 vector were inoculated into liquid SD–Ura medium for overnight culture. After normalization of $D_{660\text{ nm}}$ to 0.4, the cells were spotted in tenfold serial dilutions on SG–Ura and SD–Ura plates. The plates were incubated at 30 °C for three days before the pictures were taken. SD–Ura medium contains 2% (w/v) glucose, 0.5% (w/v) ammonium sulphate, 0.17% yeast nitrogen base with essential amino acids and other required nutrients except uracil. SG–Ura medium contains all the same components as SD–Ura except that the 2% (w/v) glucose is replaced with 2% (v/v) glycerol. For growth on agar plates, 2% agar was added.

Enzyme assays for Ndi1. The enzymatic activity of wild-type and mutant Ndi1 proteins was measured spectrophotometrically using NADH and UQ₁ as substrates. The standard assay was carried out at 30 °C in a 1.8-ml reaction mixture containing 50 mM MOPS (pH 7.0), 1 mM EDTA and varying amounts of NADH, UQ₁ and enzymes. Reactions were initiated by the addition of the enzyme to the reaction mixture. The progress of the reaction was monitored continuously by following the formation of NAD⁺ at 340 nm, using a Lambda 45 spectrophotometer (PerkinElmer Life Sciences) equipped with a magnetic stirrer in the cuvette holder. The activities were determined using an extinction coefficient for NADH of 6,220 cm⁻¹ M⁻¹ at 340 nm.

Co-immunoprecipitation experiments. Co-immunoprecipitation was performed as follows. For wild-type Ndi1, *ndi1 Δ trp1 Δ* cells were transformed with vectors pYES2–Myc–Ndi1 and pYES3–FLAG–Ndi1. After selection, the cells were grown in 50 ml of SGE–Ura medium to exponential phase before the cells were lysed by French press with extraction buffer (PBS, 0.1% Triton X-100 and protease inhibitor cocktail (Sigma)). The lysates were centrifuged at 15,000g for 30 min at 4 °C, and the supernatants were then incubated overnight with anti-FLAG or anti-Myc antibodies (Sigma) conjugated to protein A/G agarose. The agarose beads were washed three times with extraction buffer, boiled in SDS–polyacrylamide gel electrophoresis (PAGE) loading buffer, subjected to the SDS–PAGE and analysed by western blotting. Co-immunoprecipitation experiments for the truncated forms of Ndi1 were performed similarly.

Membrane fractionation. For membrane fractionation assay, *ndi1 Δ* cells transformed with pYES2–Ndi1 vectors were cultured overnight in SD–Ura medium and then transferred to SGE–Ura. After growing to the exponential phase, cells were collected by centrifugation. The cells were broken in lysis buffer (PBS buffer with protease inhibitors), and after centrifugation at 4,300g for 15 min the supernatants were obtained as the sample of total proteins. The solution of total proteins was then further ultracentrifuged at 250,000g for 1 h, and the resulting supernatant and pellet taken as samples of supernatant and membrane fractions, respectively. The total protein, supernatant and membrane fractions were subjected to western blotting.

EPR characterization of ubisemiquinone radical. Four different samples prepared from Ndi1 proteins extracted using Triton X-100 (Ndi1–TRT) or DDM (Ndi1–DDM) were tested: NADH–UQ₁–Ndi1–DDM, NADH–Ndi1–DDM, NADH–Ndi1–TRT and UQ₁–Ndi1–DDM. All EPR samples were prepared in anaerobic conditions after purification of the Ndi1 protein. The Ndi1 protein was added to a final concentration of 50 μ M and incubated with or without 5 mM UQ₁. After the addition of 2 mM NADH and incubation for 1 min, the sample was transferred to an EPR tube before a quick freezing to liquid-helium

temperature (around 2 K). EPR measurements were conducted on a Bruker X-band (9.4 GHz) EMSplus10/12 spectrometer. A cylindrical resonator (ER 4119hs TE011) was used for data collection. During EPR data acquisition, the sample temperature was maintained by an Oxford Instrument ESR 910 liquid-helium continuous-flow cryostat. A high-sensitivity EPR spectrum was obtained by temperature optimization from 2 to 200 K, and most of the EPR spectra were collected at an optimized temperature, 40 K. The EPR data acquisition parameters were as follows: modulation amplitude, 5.0 G; microwave power, 50 μ W; sweep time, 40 s; time constant, 81.92 ms; sample temperature, 40 K.

A power saturation experiment was implemented with the microwave power ranging from 44 to 14 dB (or 8 μ W to 8 mW) with a step size of 2 dB. The power saturation data was fitted following the equation $D = \sum_{i=1}^n C_i \sqrt{P(1 + P/P_{1/2,i})^{-0.5b_i}}$, in which C_i is the relative concentration of the i th free radical, $P_{1/2,i}$ is the half-saturation parameter and b_i is the homogeneity parameter. All the EPR data were collected at the High Magnetic Field Laboratory of Chinese Academy of Sciences.

31. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
32. Hall, T. A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **41**, 95–98 (1999).
33. Guindon, S. E. & Gascuel, O. A. Simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**, 696–704 (2003).
34. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
35. Collaborative Computational Project. Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
36. Schneider, T. R. & Sheldrick, G. M. Substructure solution with SHELXD. *Acta Crystallogr. D* **58**, 1772–1779 (2002).
37. Terwilliger, T. C. Rapid automatic NCS identification using heavy-atom substructures. *Acta Crystallogr. D* **58**, 2213–2215 (2002).
38. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
39. Cowtan, K. DM: an automated procedure for phase improvement by density modification. *CCP4 Newsltt.* **31**, 34–38 (1994).
40. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D* **62**, 1002–1011 (2006).
41. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
42. Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58**, 1948–1954 (2002).

CAREERS

PARTNERSHIPS UK government investment yields academic–research hires **p.485**

POSTGRADUATES International enrolment in US rises, but science gains modest **p.485**

NATUREJOBS For the latest career listings and advice www.naturejobs.com

SHUTTERSTOCK/MONGOLKA



IMMIGRATION

Waiting for green

A US ‘green-card’ visa can open up career possibilities. But getting one requires stamina — and a dash of luck.

BY KAREN KAPLAN

Genomics researcher Mark Eppinger was starting to crave stability. Since arriving in the United States in 2004 from his native Germany, Eppinger had completed a postdoc and a stint in industry, and had been working for five years at the University of Maryland in Baltimore in a non-tenure-track post. Newly married and with a baby, he wanted to find a permanent job, ideally a tenure-track academic position, and remain in the country.

But Eppinger was in his fifth year of an H-1B non-immigrant work visa, which expired in a year and could not be renewed — it has a maximum term of six years. Because he hoped to remain in the country indefinitely, Eppinger decided to apply for an immigrant visa that grants permanent-resident status in the United States to non-US citizens — the sought-after ‘green card’, so called because it used to be

issued as an actual green card. After consulting colleagues who had successfully navigated the process, Eppinger applied for a self-sponsored green card in October 2011 to US Citizenship and Immigration Services (USCIS), the federal agency that adjudicates applications and issues immigrant visas. With the help of an immigration attorney, he filed his application under two of five categories available to those who are applying for an employment-based immigration visa (see ‘Employer- or self-sponsorship’).

Although there were no glitches and he received his green card with time to spare, Eppinger found the process stressful and time-consuming. The two categories under which he had applied — as an individual of ‘extraordinary ability’ (‘Priority Worker’), and as an individual seeking a National Interest Waiver — each required substantial documentation of his professional achievements. In addition, he had to round up recommendation letters from half a dozen eminent researchers

in his field. “That was the most difficult part,” Eppinger recalls. He pre-wrote each letter, focusing on different accomplishments, then waited to get them back. “It took so long,” Eppinger says. “Even though the letter’s already written, you have to let people freely edit it.” Despite the anxiety, however, he is relieved to have the card. “You are way more flexible,” says Eppinger, now an assistant professor in microbial genomics at the University of Texas at San Antonio, a tenure-track position. “My wife and I knew we wanted to stay here — we have no intentions of going back to Germany — so going for the green card was a no-brainer.”

There are a daunting number of steps and a lot of documentation required to get a green card. But applicants can smooth their path by giving themselves ample time to understand the process, gather the necessary evidence, file paperwork and wait for application backlogs to clear.

The green card is highly sought after because researchers in the United States who hold any of the temporary, non-immigrant visas are limited in various ways. Waiting times for getting many of them can be six months or more (see *Nature* **460**, 131–132; 2009). There is the ‘home-residence’ rule that requires some visa holders to return to their home country for two years after the visa expires. Some visas limit travel in and out of the United States; others allow the visa holder to work only for the employer that sponsored the visa. And temporary visas have set terms that vary from one to six years.

DECK OF CARDS

As a permanent resident with a green card, Eppinger is now free from such worries. Green cards need to be renewed only once every 10 years and there are no limits on the number of renewals. And green-card holders can travel out of the United States for up to 179 days each year — a boon for researchers planning to attend foreign conferences or take part in international collaborations. A huge bonus for many is that having a green card confers the right to apply for all federal grants and fellowships, whereas those on non-immigrant visas are not eligible for some types of funding.

“If you’re on a self-sponsored green card,” says immigration attorney Elizabeth Goss of Tocci, Goss & Lee in Boston, Massachusetts, “you can work for whomever you want or start your own company. You’re not dependent on any employer to sponsor you for a visa. You don’t have to be beholden to anyone.”

Green-card applicants can pursue an ►

GREEN-CARD CATEGORIES

Employer- or self-sponsorship?

There are five 'preference' categories for employment-based applications (see go.nature.com/v9toa4).

● **EB-1** Priority workers, those with 'extraordinary abilities'. A select few researchers, mainly those well established in their careers, self-sponsor under this category. No labour certification required.

● **EB-2** Professionals with advanced degrees. There are three subgroups: *Exceptional ability*. Applicant must be able to show exceptional ability in the sciences and meet three out of seven criteria. But the applicant must first obtain certification from the US Department of Labor, verifying that there are insufficient qualified and willing US workers to fill the position at the prevailing wage. Employer-sponsored. *Advanced degree*. Applicant's position must

require an advanced degree and applicant must be able to prove that they have at least five years of post-degree experience in their speciality. Applicants require a labour certification. Employer-sponsored.

National Interest Waiver. Applicants request that the labour certification be waived because their working in the United States would be in the national interest. Most self-sponsoring researchers apply in this subgroup.

● **EB-3** Professionals, skilled workers and unskilled workers.

● **EB-4** Certain immigrants, including those who have served or are serving abroad for the US federal government.

● **EB-5** Immigrants who are investing at least US\$500,000 in a new commercial business that will create US jobs. **K.K.**

► employment-based visa, which requires a certain level of skills, education and/or work experience, or a family-based visa, which requires sponsorship by a spouse or relative who is a US citizen or legal permanent resident. Once a researcher decides to apply for a green card, the next determination is the type of sponsorship to pursue. For employment-based applications, the options are employer sponsorship or self-sponsorship. Employer sponsorship requires a job offer or an existing position; in both cases, the applicant must expect to remain in that job for up to two years. If they leave that job within six months of getting the card, it may be revoked.

There are other restrictions to employer sponsorship: employers may choose to sponsor only particular positions, or only employees who will be in a particular position for a certain length of time. And employers may require an employee to work for some time before he or she becomes eligible for sponsorship. However, the employer generally organizes the process and must pay the processing fees and other expenses of the application, such as the cost of the required medical examination.

Many scientists, however, choose self-sponsorship, which does not depend on having a job or job offer. Immigration experts note other benefits: a self-sponsoring applicant can soon move to another job — as long as it's in the same field — or launch his or her own company.

The application steps for employer and self-sponsorship are very similar (see 'Securing the visa'). But anyone applying under the Priority Worker category (employment-based category 1, or EB-1) must prove that they have "extraordinary" abilities and are internationally renowned in their field. A doctorate or postdoc experience at an elite US university

will not be sufficient. "You've got to show that you're one of the small percentage at the very top of your field," says immigration attorney Brendan Delaney of Leavy, Frank & Delaney in Bethesda, Maryland, who, as a consultant for the US National Postdoctoral Association in Washington DC, works with many early-career researchers.

As a result, most early-career researchers opt for the National Interest Waiver subcategory of professionals with advanced degrees

(EB-2). The other stages are as for an employer-sponsored application, but the applicant has the responsibility of providing all the information needed and of monitoring their application. After submitting an I-140, they receive their 'priority date' — their place in line for an immigrant visa number. Once the priority date is current and the I-140 is granted, the applicant must then file an I-485 — the final stage.

Evidence submitted at the I-140 stage of a self-sponsored application could include copies of published papers along with information on first or second authorship, the papers' citation indexes and the impact factor of the journals in which they appeared; records of interviews and conference talks, especially invited talks; peer review undertaken; patents filed, especially if in use or generating interest; and awards, grants and fellowships (see 'Gathering the evidence').

BEWARE THE CAVEATS

All these documents — along with a published book and grants in hand — were what it took for Dimitar Baronov, a Bulgarian native with a PhD from Boston University in Massachusetts, to secure a self-sponsored green card. The card was key to Baronov's plans, and in 2010 he co-founded Sterling Point Research in Boston, Massachusetts, which develops medical diagnostics. Baronov says that he has far more flexibility. "It would have been really hard to run this company if I were still a temporary-visa holder," he says. "I can apply for federal grants, I have latitude for my research and I can create jobs."

But applications don't come cheap. Self-sponsoring applicants who hire an immigration attorney — and immigration experts

STEP-BY-STEP

Securing the visa

Typically, there are four or five stages to a green-card application, whether employer-sponsored or self-sponsored.

● The employer or self-sponsor must determine which of five employment-based (EB) immigrant visa categories under which to petition (see 'Employer- or self-sponsorship').

● For employer-sponsored green cards, the employer must get labour certification from the US Department of Labor in some cases. It is not necessary under any subgroups of EB1 ('Priority Worker'), or the National Interest Waiver subgroup of EB2. Certification can take up to eight months.

● Next, the employer or self-sponsor must file an Immigrant Petition for Alien Worker, known as an I-140, to the US Citizenship and Immigration Services (USCIS), which adjudicates the application and, on approval, sends the employer a notice with the applicant's 'priority date' — their place

in line for an immigrant visa number. The agency will also send the approved petition to the US Department of State's National Visa Center, where it will remain until an immigrant visa number becomes available.

● The sponsor must then check the state department's monthly Visa Bulletin to determine whether the priority date is backlogged. If it is — because the applicant's country of origin is oversubscribed — he or she may have to wait for months or years. Each country has an annual quota, so applicants from oversubscribed countries, such as Mexico, China, India and the Philippines, can face years of delay.

● Once the priority date is current, the sponsor files an Application to Adjust Status (also known as an Application to Register Permanent Residence), called the I-485, to the USCIS — and the applicant gets their green card. **K.K.**

recommend doing that — will rack up legal and federal processing fees and other expenses of up to US\$8,000.

There are other caveats. One is that green-card renewal, although usually routine, requires the holder to be free of certain criminal convictions. Some missteps, such as overstaying a previous visa, may render the applicant ineligible. Another sticking point is that trips out of the country beyond 179 days require USCIS re-entry permits, which provide for absences of up to two years. Scientists who travel outside of the country on collaborations or fieldwork for more than six months at a time should be aware that the USCIS might interpret those trips as abandonment of their US residence and move to deport them.

Opinions vary on the best time for an early-career researcher to pursue a green card. Some immigration experts say that it can take as long as eight years after a researcher's US career launch, depending on the quality of their publication and citation records.

With so much riding on the card, a rush to apply can create anxiety. Hou-Sung Jung, a research assistant professor of plant biology at Dartmouth College in Hanover, New Hampshire, experienced this first-hand. The South Korea native spent seven years as a post-doc, mostly on an H-1B visa. Busy with his research, publications and a new marriage, he forgot about the visa's expiry date. He was able to assemble the documentation and materials for a green-card application within two months and received the card with a month to spare. "I was lucky," says Jung.

Some immigration experts counsel that a researcher arriving in the United States on a temporary visa of any type should immediately start educating themselves on the process and details of acquiring the green card, even if they don't expect to want one. "If you leave things too late, you may find yourself in a position where you're running out of options," says

Delaney, who, as a native of Northern Ireland, has gone through the process himself.

Applicants may receive conflicting or inaccurate information from employers that steers them down the wrong path. "At the green-card centre at one university, the information on the process was so confusing," says Sadakatsu Ikeda, a native of Japan who is a clinical fellow

TOP TIPS

Gathering the evidence

- Start amassing necessary documentation as early as possible: birth certificate, academic degrees, proof of scientific-society and association memberships, invited conference presentations, book chapters, copies of or links to media interviews and their dates, awards and dates, published manuscripts, citation indexes, journal impact factors.
- Network at every chance you get: you need to create an association of high-level colleagues outside your institution with whom you have not collaborated, including, for example, contacts from conferences or talks. They will be your sources for up to five letters of recommendation, which will be more influential than those from institutional or collaborative colleagues.
- Write the letters yourself, focusing on different accomplishments for each that illustrate how you fit into the category under which you're applying. Send to the appropriate recipient with a stamped, self-addressed envelope and provide a deadline. **K.K.**

in haematology and oncology at the University of Michigan in Ann Arbor. "Different people told me different things." Ultimately, he decided to pursue the self-sponsored route for the freedom to move to different employers and to control the process himself.

Shortly after starting a molecular biology and genetics postdoc at Harvard Medical School in Boston, Massachusetts, Ikeda hired an attorney who showed him how to apply under the National Interest Waiver subgroup and recommended that Ikeda work on publishing citable papers, building an extensive network of contacts and getting his name widely known in his field. By the time Ikeda began to pursue the application seven years later, after a three-year residency, he was first or second author on a number of papers in high-impact journals, had amassed a robust contacts network and was well known in cancer research. He had no trouble getting five letters of recommendation, and received his green card without any hassle six months later.

"Getting the right information was a big struggle in the very beginning, and building this network and my reputation — that was the biggest obstacle," says Ikeda. "But in the end, I was successful because I had a good roadmap." ■

Karen Kaplan is assistant Careers editor at Nature.

PARTNERSHIPS

UK recruitment drive

Recruitment of academic researchers is under way at Queen's University Belfast (QUB), the University of Manchester and University College London (UCL) in connection with around £1 billion (US\$1.6 billion) in awards from the UK government's Research Partnership Investment Fund and associated philanthropic donations. QUB will use its £32-million share to help hire 500 researchers and clinicians for its health-sciences institute, which includes a new research centre for eye disease and diabetes slated for completion in 2016. Manchester will hire about 20 senior academics in 2013 for cancer research, and UCL expects to create more posts at a new centre for rare paediatric diseases scheduled to open in 2018.

EDUCATION

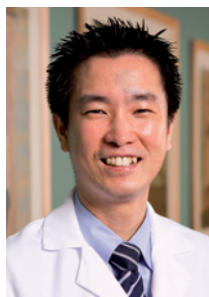
Teachers without PhDs

British postgraduate students might often be better qualified than their lecturers, according to a study, *Academic Staff in UK Higher Education Institutions*, released on 1 November. The report finds that just over half of the full-time academic staff across all disciplines and institutions have doctoral degrees. Author Malcolm Tight at Lancaster University, UK, warns that the results reflect poorly on the quality of postgraduate education at some UK universities. He argues that universities have focused on diversifying their student base at the expense of developing the quality of their faculty. "It's certainly possible that a doctoral student might have more knowledge of a discipline than her or his professor," Tight says.

POSTGRADUATES

International enrolment

First-time enrolment of international graduate students in the United States rose 8% for 2011–12 across all academic fields, matching the previous year's increase, according to the latest survey by the US Council of Graduate Schools in Washington DC. Gains in the sciences alone were more modest, with life sciences up 1% and physical sciences 4%. Business and engineering led all fields with gains of 15% and 12% respectively. China continues to send increasing numbers of graduate students to the US; enrolments were up 22% from the previous year, the seventh consecutive year of growth.



"Getting the right information was a big struggle in the very beginning."

Sadakatsu Ikeda

LET SLIP THE DOGS

The art of soldiering on.

BY WILLIAM T. VANDEMARK

In Abuja, Nigeria, fresh from an open-air market, a woman dressed in black walks with a melon balanced on her head.

"Yeah, baby. William Tell time," says Specialist Browning, a Marine bored out of his skull. He and his squad have been on duty 100 hours straight, hopped up on orexin-z to counter sleep deprivation.

"Don't even," says Sergeant Knox, but Browning's beam rifle sings, melon splatters, and rind flies like shrapnel. The woman stands in disbelief, arms spread, while juice — thin as artificial blood — drips from her abaya.

"Huzzah," says Browning. The woman fixes his position. She gestures wildly and advances.

Knox steps from the checkpoint's ceramic-plated kiosk. "Lady, stop right there," he shouts.

A small crowd gathers. Someone tosses a bottle. As it arcs through the air, an AI built into an auto-turret atop the kiosk assesses the risk. The bottle shatters in the street.

Knox unholsters his pistol and waves it at the woman, trying to dissuade her from entering the checkpoint's no-go zone — an area mined with cybernetic fleas. He hates having to fill out reports when civvies need to be deloused. But without pause, the woman steps into the orange dust, and it comes alive. She stomps, swats, screams in frustration. Stepping back, she points at Knox as if he is the one to blame and reaches into her sleeve.

Someone flings a tin can at the Marines. A magnetometer chatters as if it's a grenade. The auto-turret sizzles, spitting streams of nano flechettes at hypersonic velocities. The can and its contents vaporize; the superheated air rebounds with a thunderclap.

Knox flinches; his pistol discharges.

In sudden silence, the woman drops to her knees and slumps forward until her forehead touches the ground. But this isn't sajdah, a prostration to her God; the back of her skull is gone. Tangled hair, flaps of skin and remnants of her hijab drape the exit wound like a shroud. In an outstretched, angled hand, she clutches a melon rind.

The cyberfleas ignore the fruit. They gorge on blood.

Deep within a hospital ship, off the coast of Nigeria, Sergeant Knox lies on a polyglass table, head encircled by a halo. A psych-tech asks for specifics. She wants a description of the woman, of the market, of the three-



legged dog Knox saw earlier that day; she wants childhood memories of pets; she wants Rorschach associations of the woman's blood spatter pattern.

As she works through Knox's story again and again, teasing new details from his narrative, she walks about an enlarged, holographic projection of his brain. Within the stereotactic model, neural pathways flicker and fluorescent voxels coalesce: a memory trace.

"You're on a hippocampus block," the tech says. "So you won't remember any of this, but informed consent requires me to explain that when we strip the problematic engram, secondary pathways may be affected. Although the goal is event-removal, phantom reconstruction will occur. The mind heals by filling in gaps. But really, you'll be fine. And tomorrow, you'll wake up on R&R, unaware of the synaptic pruning. No Post Combat Stress Disorder for this Marine. Right, Sergeant?"

In Abuja, Nigeria, fresh from an open-air market, a woman dressed in black walks with a melon balanced on her head.

"Oh, baby. I want a taste of that," says Specialist Browning, a Marine bored out of his skull. He and his squad have been on duty 100 hours straight, hopped up on orexin-z.

"Stand down," says Sergeant Knox, but Browning lets loose with a wolf-whistle.

Before the sound dies, a three-legged dog

lopes from a shaded doorway. Knox recognizes it as the stray he'd fed while on patrol. He laughs. "Yeah, Browning. You sure can attract the ladies."

At Knox's voice, the dog makes a beeline for the kiosk and heads straight into the no-go zone.

"Git!" Knox shouts.

But the dust awakens. The dog yips, bites at its legs, snaps at the air. It spins about, activating even more of the cybernetic fleas.

The woman with the melon gestures at the dog and yells at Knox in a language he doesn't understand, but her message is all too clear.

"No," Knox says. "Not my fault. You want him? You get him." Trying to wrangle a feral dog, near rabid with pain, isn't Knox's idea of galantry. Besides, he's allergic to the fleas' caustic discharge.

As if to plead its case, the dog howls, its distended voice raw and unearthly.

Jesus, Knox thinks. Why doesn't it just run away?

The bustle of the market has stopped. Everyone is watching. Even the kiosk's auto-turret seems transfixed, its barrel jerking in micro adjustments as its AI rips through algorithms. Assessing. Assessing. Assessing.

But it won't pull the trigger. Of course not. Such dirty work always falls to Knox. Cursing, he unholsters his pistol and takes aim...

The dog is dead before it hits the ground. It's disconcerting the way its limbs splay. Almost humanlike.

Knox glances at his pistol. Although smoke is curling from its barrel, he doesn't recall pulling the trigger. He looks up. His own men, the people in the market... they're staring at him. He catches sight of the woman and for an instant it seems as if her face is alive with bugs.

The world cants.

Knox drops to his knees and rocks back and forth as waves of guilt wash over him.

Keep it together, he tells himself. It was just a dog. Just a dog.

He closes his eyes, leans forward and rests his forehead on the ground, wishing to God none of it were real. ■

William T. Vandemark *chases storms, photographs weather vanes and prospects for fulgurites. His work has appeared in Apex Magazine, Intergalactic Medicine Show and assorted anthologies.*

NATURE.COM
Follow Futures:
@NatureFutures
go.nature.com/mtoodm